

AN ABSTRACT OF THE DISSERTATION OF

John B. Henry III for the degree of Doctor of Philosophy in Statistics presented on
May 27, 2008.

Title:

Extreme Value Index Estimation with Applications to Modeling Extreme Insurance
Losses and Sea Surface Temperatures

Abstract approved: _____

Mina E Ossiander

The extreme value index (EVI) links the generalized extreme value (GEV) distribution and the generalized Pareto (GP) distribution. These two distributions are fundamental in extreme value theory (EVT), with the GEV distribution being the only possible non-degenerate limiting distribution of properly normalized maxima of iid random variables, and the GP distribution appearing as the limit distribution of scaled excesses over high thresholds. The reciprocal of the EVI is known as the Pareto tail index (provided the EVI is positive). A new tail index estimator is proposed here that is obtained by matching general theoretical harmonic moments with corresponding empirical moments. Theoretical properties of the estimator are provided along with applications and a comparison to other tail index estimators. A tail index estimator suitable for partitioned data is also given. Having only partitioned data to work with is a circumstance sometimes faced by actuaries. Strengths and weaknesses of this estimator are explored through simulation. An application of the estimator to real world partitioned insurance data is given. The sign of the EVI is of interest when one is interested in testing whether or not a distribution

has a bounded tail. In this work, the controversial thermostat hypothesis for sea surface temperature (SST) is investigated in this framework. A GP model using SST data from the western Pacific warm pool provides some evidence of a temperature upper bound between 31.2°C and 32.0°C . This estimate is compared those obtained elsewhere using the ocean heat budget and other physical models.

©Copyright by John B. Henry III

May 27, 2008

All Rights Reserved

Extreme Value Index Estimation with Applications to Modeling Extreme Insurance
Losses and Sea Surface Temperatures

by

John B. Henry III

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Presented May 27, 2008
Commencement June 2009

Doctor of Philosophy dissertation of John B. Henry III presented on May 27, 2008

APPROVED:

Major Professor, representing Statistics

Chair of the Department of Statistics

Dean of the Graduate School

I understand that my dissertation will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my thesis to any reader upon request.

John B. Henry III, Author

ACKNOWLEDGEMENTS

Thank you to Mina Ossiander for her guidance during my graduate program and for her helpful comments and suggestions throughout the writing of this dissertation. Thanks you also to Ping-Hung Hsieh for introducing me to extreme value theory and working with me on my first paper. Thank you to Ted Strub for his patience and advise, and to Lisa Madsen, Alix Gitelman, Yevgeniy Kovchegov, and Kevin Boston for being on my PhD committee.

Thank you to the Statistics Department at Oregon State University, especially those who taught my graduate courses in Statistics: David Birkes, Robert Smythe, Daniel Schafer, Paul Murtaugh, Virginia Lesser, Alix Gitelman, Jeffrey Arther, and Annie Que. A special thank you to David Birkes for all of his help during midnight office hours.

Thank you to Julia Jones, Katherine Hoffman, and the IGERT Ecosystem Informatics program at Oregon State University. Being part of this program was beneficial to me in so many ways. Thank you to Rick Katz and Joanie Kleypas at the National Center for Atmospheric Research for their time and assistance during my IGERT internship.

Thank you to my wonderful wife Jodie. I am truly grateful for her support, encouragement, and hard work. Thank you to my beautiful kids John and Ella! Thank you to Chayah and Josh, Wally and Kristen, Lewis, Nick and Deeanne, Randy and Carleen, Sara and Murry, and all of the rest of my family for the ways that they have helped me.

Thank you Nan.

CONTRIBUTION OF AUTHORS

Ping-Hung Hsieh is a co-author of the original paper that is Chapter 3 of this dissertation.

DEDICATION

To Lil John and Ella Joy

TABLE OF CONTENTS

| | <u>Page</u> |
|---|-------------|
| 1. INTRODUCTION | 1 |
| 1.1. EVT Background..... | 1 |
| 1.2. Statement of Problems and Organization of this Dissertation..... | 2 |
| 2. A HARMONIC MOMENT TAIL INDEX ESTIMATOR | 4 |
| 2.1. Introduction..... | 5 |
| 2.2. The Harmonic Moment Estimator | 8 |
| 2.2.1 Efficiency of the Harmonic Moment Estimator..... | 10 |
| 2.2.2 Robustness of the Harmonic Moment Estimator | 12 |
| 2.2.3 Investigating the Choice of θ | 15 |
| 2.3. Simulation Study | 17 |
| 2.3.1 Description | 17 |
| 2.3.2 Results | 18 |
| 2.4. Threshold Selection: A new Sum Plot | 21 |
| 2.5. Applications in Insurance..... | 23 |
| 2.6. Conclusion | 26 |

TABLE OF CONTENTS (Continued)

| | <u>Page</u> |
|--|-------------|
| 3. EXTREME VALUE ANALYSIS FOR PARTITIONED INSURANCE LOSSES | 28 |
| 3.1. Introduction | 29 |
| 3.2. Literature Review | 31 |
| 3.2.1 The Hill and Pickands Estimators | 32 |
| 3.2.2 Some Recent Tail Index Estimators | 34 |
| 3.2.2.1 Censored Data Estimator | 34 |
| 3.2.2.2 Location Invariant Estimators | 35 |
| 3.2.2.3 Generalized Median Estimator | 36 |
| 3.2.2.4 Probability Integral Transform Statistic Estimator | 37 |
| 3.3. Tail Index Estimator for Partitioned Data | 38 |
| 3.4. Performance Assessment | 40 |
| 3.5. Discussion on Simulation Results | 45 |
| 3.6. Applications to Insurance | 50 |
| 3.7. Summary and Conclusion | 53 |
| 4. MODELING SEA SURFACE TEMPERATURE EXTREMES: A LOOK AT THE THERMOSTAT HYPOTHESIS | 55 |

TABLE OF CONTENTS (Continued)

| | <u>Page</u> |
|--|-------------|
| 4.1. Introduction..... | 56 |
| 4.1.1 The Thermostat Hypothesis..... | 56 |
| 4.1.2 WPWP SST Data | 58 |
| 4.1.3 Organization of paper | 59 |
| 4.2. EVT Background..... | 59 |
| 4.2.1 Method of Block Maxima | 59 |
| 4.2.1.1 Parameter Estimation | 62 |
| 4.2.2 Threshold Models..... | 62 |
| 4.2.2.1 Parameter Estimation | 65 |
| 4.2.2.2 Threshold Selection..... | 65 |
| 4.3. Methods..... | 66 |
| 4.3.1 SST Extremes: Trend and Seasonality..... | 66 |
| 4.3.2 Residuals and Excess | 69 |
| 4.3.3 GP Model for SST | 70 |
| 4.4. Results..... | 72 |
| 4.4.1 Trend and Seasonality | 72 |
| 4.4.2 Estimated Upper Bound | 73 |

TABLE OF CONTENTS (Continued)

| | <u>Page</u> |
|-------------------------------------|-------------|
| 4.5. Discussion and Conclusion..... | 76 |
| 5. GENERAL CONCLUSIONS | 78 |
| BIBLIOGRAPHY | 80 |

LIST OF FIGURES

| <u>Figure</u> | <u>Page</u> |
|---|-------------|
| 2.1 Comparison of sum plots | 22 |
| 2.2 Diagnostic plots for the Secura Belgian Re automobile claims data | 24 |
| 2.3 Pareto fit for the Secura Belgian Re automobile claims | 26 |
| 3.1 Loss in efficiency due to partitioned data | 43 |
| 3.2 Robustness of Hill and G_k | 44 |
| 3.3 Tail index estimation for fire loss data | 52 |
| 3.4 Empirical and model fits for fire loss data | 53 |
| 4.1 GEV Densities | 61 |
| 4.2 GP Densities | 63 |
| 4.3 SST Trends | 67 |
| 4.4 SST Seasonality | 68 |
| 4.5 SST residuals | 69 |
| 4.6 Sampling distributions of $\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{\xi}$ | 72 |
| 4.7 GPD fit to SST WPWP data | 74 |
| 4.8 Estimated SST upper bound | 75 |

LIST OF TABLES

| <u>Table</u> | <u>Page</u> |
|--|-------------|
| 2.1 Comparison of $\text{HM}_{k,n}(\theta)$ with other estimators: Bias, SD, and MSE | 19 |
| 2.2 Estimates of $\Pi(R)$ | 25 |
| 3.1 Tail index parameters and mean excess functions for selected distributions | 41 |
| 3.2 Loss of efficiency with the use of partitioned data | 46 |
| 3.3 Robustness of G_k | 47 |
| 3.4 Homeowners physical damage grouped data | 50 |
| 4.1 Fitted SST trend and seasonal components | 73 |
| 4.2 Estimated limiting SST upper bound | 76 |

EXTREME VALUE INDEX ESTIMATION WITH APPLICATIONS TO MODELING EXTREME INSURANCE LOSSES AND SEA SURFACE TEMPERATURES

1. INTRODUCTION

1.1. EVT Background

Extreme value theory (EVT) is a branch of probability theory that provides methods for modeling and analyzing extreme phenomena. Some of the mathematical tools used include point processes, and regular variation. Statistical inference is based on point estimation, hypothesis testing, simulation, and model diagnostics. Applications of EVT are found in many areas including hydrology, meteorology, geology, and seismic analysis. A few examples of problems relying on EVT include:

- 100 year flood level estimation
- *value-at-risk* determination in finance
- premium for excess of loss over high retention levels reinsurance treaties

EVT is a developing area, which over the last two decades has received a tremendous amount of attention in the literature. The roots of the theory, however, began with the

early work of M. Fréchet (1927), R. Fisher and L. Tippett (1928), and R. von Mises (1936), and B. Gnedenko (1943). This early work was focused on the possible (non-degenerate) limiting distributions of the maxima of sequences of iid random variables. The statistical theory was initiated by J. Pickands III (1975) and B. Hill (1975).

1.2. Statement of Problems and Organization of this Dissertation

Excess over thresholds play a fundamental role in many fields, and EVT suggests using the generalized Pareto (GP) distribution with distribution function (d.f.)

$$H_{\xi,\sigma}(y) = \begin{cases} 1 - \left(1 + \frac{\xi y}{\sigma}\right)^{-1/\xi} & \text{if } \xi \neq 0 \\ 1 - \exp\{-y/\sigma\} & \text{if } \xi = 0 \end{cases} \quad (1.1)$$

where $y \in [0, \infty)$ when $\xi \geq 0$ and $y \in [0, -\sigma/\xi]$ when $\xi < 0$, to model excess over thresholds. The shape parameter ξ , known as the extreme value index (EVI), characterized the tail of H , which can be exponential ($\xi = 0$), bounded ($\xi < 0$), or power law ($\xi > 0$). A tremendous amount of work is seen in the literature devoted to estimation methods for ξ and applications.

In the case that $\xi > 0$, we have $1 - H_{\xi,\sigma}(y) \sim Cy^{-\alpha}$ where $\alpha = 1/\xi$ and \sim indicates the limit of the ratio of the two functions is equal to 1 as $y \rightarrow \infty$. The term $Cy^{-\alpha}$ is easily recognized as the Pareto survivor function. The problem of how to estimate α in this case is crucial in many applications. In Chapter 2 a new estimator for the tail index parameter is proposed that is obtained by matching theoretical and empirical harmonic moments

assuming a Pareto distribution above a random threshold. Asymptotic and robustness properties of the estimator are worked out and the estimator is compared to others in a simulation study. Applications to modeling insurance losses are also provided.

In Chapter 3 a new tail index estimator appropriate for grouped data is developed with a focus on actuarial applications. The question of how to estimate α in the case that $\xi > 0$ provides additional challenges if one is working with partitioned data. This is useful because actuaries are sometimes required to work with grouped loss data, and must still be able to make inference about the tail of the loss distribution.

In Chapter 4 the thermostat hypothesis for sea surface temperatures is discussed and is looked at from an extreme value theory point of view. In particular, the thermostat hypothesis is looked at as a question as to whether or not sea surface temperatures have a distribution with a bounded tail. Recall that the tail of H in (1.1) is bounded in the case that $\xi < 0$. Whether such a hypothesis holds is of interest to marine ecologists and has important implications for the potential impacts of global warming on coral reefs.

2. A HARMONIC MOMENT TAIL INDEX ESTIMATOR

John B. Henry III

Abstract

A new tail index estimator is proposed that is obtained by matching general theoretical harmonic moments with corresponding empirical moments. The estimator is shown to be robust and works particularly well in small sample situations. A tuning parameter θ is included that allows a tradeoff between efficiency and robustness of the estimator to be made. The choice of θ that makes the estimator optimal in a mean squared error sense is derived. A simulation study compares the estimator to other tail index estimators, including the well known Hill estimator, which is shown to be a special limiting case of the proposed estimator. Examples are provided using insurance loss data and a new threshold selection approach is included.

2.1. Introduction

Estimation of the Pareto tail index parameter is important in many applications including insurance, finance, geology, and hydrology. See, for example, Embrechts et al (1997), Caers et al (1999), and Katz et al (2002). There have been many tail index estimators proposed since the early work of Hill (1975) and Pickands (1975). In this decade alone, some of the estimators that have been proposed can be found in Brazauskas and Serfling (2000), Gomes et al (2000), Fraga Alves (2001), Peng and Welsh (2001), Hsieh (2002), Groeneboom et al (2003), Juarez and Schucany (2004), Seagers (2005), and Finkelstein et al (2006). These estimators are typically functions of the largest k out of n order statistics. Some recent estimators proposed by Bianchi and Meerschaert (2000), Frigessi et al (2002), and Fialová et al (2004), make use of all order statistics. There are also tail index estimators for special cases such as when one is working with censored data (Beirlant and Guillou (2001) and Beirlant et al (2007)) or grouped data (Henry III and Hsieh (submitted)). A summary of many existing tail index estimators and their properties can be found in Brazauskas and Serfling (2000), Hsieh (2002) Beirlant et al (2004), Markovich (2007), and Gomes et al (2008).

The estimator proposed in this work is applicable for distributions F in the maximum domain of attraction of the Fréchet distribution. That is,

$$\bar{F}(x) \equiv 1 - F(x) = \ell(x)x^{-\alpha}, \text{ for some } \alpha > 0, \quad (2.1)$$

where $\ell(x)$ is slowly varying at infinity. Note that $\ell(x)$ satisfies $\ell(tx) \sim \ell(x)$ for any

$t > 0$, with \sim indicating the limit of the ratio of the two functions is 1 as $x \rightarrow \infty$. Many commonly used distributions satisfy (2.1) including the Pareto, Burr, Fréchet, half T, F, inverse gamma, and log gamma. It is worth noting here that distributions satisfying (2.1) can be closely approximated by an exact Pareto distribution above high thresholds. In particular, suppose X is a r.v. with d.f. F satisfying (2.1) and consider $Y = X|X > u$ for large u . Let G denote the distribution function of Y , $y = tu$, $t > 1$ fixed and $u \in \mathbb{R}^+$. Then

$$\bar{G}(y) = \frac{\bar{F}(tu)}{\bar{F}(u)} = \frac{\ell(tu)}{\ell(u)} t^{-\alpha} \sim cy^{-\alpha}, \quad u \rightarrow \infty, \text{ where } c = u^\alpha.$$

Hence, for large enough x ,

$$\bar{F}(x) = \bar{G}(x)\bar{F}(u) \approx cx^{-\alpha}, \quad \text{where } c = \bar{F}(u)u^\alpha.$$

Note that the power law on the right hand side is the Pareto tail probability function. In the development of the harmonic moment estimator below, it is assumed that there is an exact Pareto tail beyond some high threshold u . That is,

$$\exists u > 0 \quad \text{such that} \quad x > u \Rightarrow P(X > x) = cx^{-\alpha}, \quad c, \alpha > 0. \quad (2.2)$$

The tail index parameter α is a measure of how heavy-tailed the heavy tailed distribution is. In fact, $E(X^k)$ is infinite for $\alpha < k$. Examples of estimates for α are often observed in insurance applications in the interval $(0, 2)$. See, for example, McNeil (1997), and Hsieh (1999).

Properties of the estimator being proposed here are derived in Section 2.2., and it is

compared to the well known Hill estimator (Hill (1975)) given by

$$\text{Hill}_{k,n} = \left\{ \frac{1}{k} \sum_{i=1}^k \log \left(\frac{X^{(i)}}{X^{(k+1)}} \right) \right\}^{-1}, \quad k \in \{1, \dots, n-1\}, \quad (2.3)$$

where X_1, \dots, X_n is a sequence of independent copies of X and $X^{(1)} \geq \dots \geq X^{(n)}$ denote the descending order statistics. The Hill estimator is a conditional maximum likelihood estimator (MLE) and it can be shown that $\text{Hill}_{k,n} \xrightarrow{P} \alpha$ and $\sqrt{k}(\text{Hill}_{k,n} - \alpha) \xrightarrow{d} N(0, \alpha^2)$. We note that $\text{Hill}_{k,n}$ has asymptotic variance (AVar) equal to α^2/k , which is the Cramer-Rao lower bound for the variance of any unbiased estimator of α . As a result, for an estimator $\hat{\alpha}$ of α , one is often interested in the asymptotic relative efficiency (AREff) of $\text{Hill}_{k,n}$ to $\hat{\alpha}$, $\lim_{n \rightarrow \infty} \text{AVar}(\text{Hill}_{k,n}) / \text{AVar}(\hat{\alpha})$. While this quantity cannot exceed 1, it is shown that the asymptotic relative efficiency of the proposed estimator can be made as close to 1 as one likes by adjusting a tuning parameter. More importantly, however, a simulation study in Section 2.3. shows that for small sample sizes the proposed estimator can outperform the Hill estimator in terms of bias, variance, and mean squared error (MSE). This result is important because one often uses only a small number of the largest order statistics when estimating the tail index.

The asymptotic contamination breakdown point (BP) of an estimator is one measure of robustness. While the Hill estimator has asymptotic BP of zero, the new tail index estimator described here has asymptotic BP in $(0, 1/2]$ depending on the true value of α and the choice of the tuning parameter θ . The gross error sensitivity (GES) is, for large k , a measure of the maximum impact that the change in a single observation has on an estimator. The GES of the proposed estimator is also given as a measure of robustness.

A discussion is given on how a trade-off between efficiency and robustness can be made by adjustment of θ . These robustness properties are derived and discussed in Section 2.2.. A more detailed discussion of BP's and GES can be found in Maronna et al (2006).

A new diagnostic plot is described in Section 2.4. that aids in determining the number k of upper order statistics to use in the estimator for α . This plot follows naturally from the estimator being proposed here, and is compared to another diagnostic plot in a few examples. Section 2.5. provides an example using the new estimator for α and diagnostic plot with insurance data, followed by concluding remarks in Section 2.6..

2.2. The Harmonic Moment Estimator

For a moment, consider the case that X has an exact Pareto(α) distribution with d.f. $F(x) = 1 - (x/\sigma)^{-\alpha}$, for $x \geq \sigma > 0$, where $\alpha > 0$ is unknown and σ is known. For extremely heavy tailed Pareto distributions ($0 < \alpha < 1$), estimators for α such as the method of moments (MOM) and probability weighted moments (PWM) estimators do not apply. One can instead consider the harmonic mean as another measure of location. The harmonic mean for a random variable X is given by $\mu_h := \{E(X^{-1})\}^{-1}$. For a Pareto(α) r.v. we have $\mu_h = \sigma(1 + \alpha^{-1})$. Then one obtains an estimator for α , for a sample for size n , by setting μ_h equal to the sample harmonic mean, $(\frac{1}{n} \sum_1^n X_i^{-1})^{-1}$, and solving for α . It is interesting to note that using the same idea with the geometric mean leads to the MLE estimator for α . Rather than looking at the first harmonic moment,

however, one can consider general harmonic moments resulting in the estimator

$$\hat{\alpha} = \frac{\frac{1}{n} \sum_{i=1}^n \left(\frac{\sigma}{X_i} \right)^{1/\theta}}{\theta \left(1 - \frac{1}{n} \sum_{i=1}^n \left(\frac{\sigma}{X_i} \right)^{1/\theta} \right)}, \quad \theta \in \mathbb{R}^+. \quad (2.4)$$

Now consider the case that X satisfies (2.2). Note that here we are assuming a Pareto tail above a threshold u , but the form of $F(x)$ is not specified for $x < u$. The above idea leads to the following proposed estimator for α using the upper order statistics above a random threshold $X^{(k+1)}$.

$$\text{HM}_{k,n}(\theta) := \frac{\frac{1}{k} \sum_{i=1}^k \left(\frac{X^{(k+1)}}{X^{(i)}} \right)^{1/\theta}}{\theta \left(1 - \frac{1}{k} \sum_{i=1}^k \left(\frac{X^{(k+1)}}{X^{(i)}} \right)^{1/\theta} \right)}, \quad \theta \in \mathbb{R}^+, \quad k \in \{1, \dots, n-1\}. \quad (2.5)$$

Modified versions of $\text{HM}_{k,n}(\theta)$ are given in Equations (2.19) and (2.20) below. It is shown later how a choice of the tuning parameter θ influences the efficiency and robustness of $\text{HM}_{k,n}(\theta)$. A robust version of $\text{HM}_{k,n}(\theta)$ is in Equation (2.12), while a minimal mean squared error (MSE) version of $\text{HM}_{k,n}(\theta)$ is given in Equation (2.18). The following lemma will be useful in developing properties of this estimator.

Lemma 2.2.0.1 *Given $X^{(k+1)} \geq u$, we have $\text{HM}_{k,n}(\theta) \stackrel{d}{=} \theta^{-1} \bar{Y}_k (1 - \bar{Y}_k)^{-1}$ where Y_1, \dots, Y_n are iid Beta random variables, each with common d.f. $F_Y(y) = y^{\alpha\theta}$, $0 \leq y \leq 1$. Here $\bar{Y}_k := \frac{1}{k} \sum_{i=1}^k Y_i$.*

Proof Using Rényi's representation theorem (Rényi 1953), it can be shown that

$$\alpha i \log \frac{X^{(i)}}{X^{(i+1)}} \stackrel{d}{=} E_i, \quad i = 1, \dots, k,$$

where the E_i 's are independent mean one exponential random variables. Notice that

$$\frac{1}{k} \sum_{i=1}^k -\log \left\{ \left(\frac{X^{(k+1)}}{X^{(i)}} \right)^{1/\theta} \right\} = \frac{1}{k} \sum_{i=1}^k (1/\theta) i \log \left(\frac{X^{(i)}}{X^{(i+1)}} \right) \stackrel{d}{=} \frac{1}{k} \sum_{i=1}^k \frac{1}{\alpha\theta} E_i.$$

As a result,

$$\frac{1}{k} \sum_{i=1}^k \left(\frac{X^{(k+1)}}{X^{(i)}} \right)^{1/\theta} \stackrel{d}{=} \frac{1}{k} \sum_{i=1}^k \exp \left\{ \frac{-E_i}{\alpha\theta} \right\}.$$

It is easy to check that $\exp \left\{ \frac{-E_i}{\alpha\theta} \right\} \stackrel{d}{=} Y_i$, $i = 1, \dots, k$ are i.i.d. with d.f. $F_Y(y) = y^{\alpha\theta}$, where $0 \leq y \leq 1$. Hence $\frac{1}{k} \sum_{i=1}^k (X^{(k+1)}/X^{(i)})^{1/\theta} \stackrel{d}{=} \bar{Y}_k$ and $\text{HM}_{k,n}(\theta) \stackrel{d}{=} \theta^{-1} \bar{Y}_k (1 - \bar{Y}_k)^{-1}$. \square

Before discussing efficiency and robustness note that $g(\bar{Y}_k) \stackrel{d}{=} \text{HM}_{k,n}(\theta)$ where $g(t) := \theta^{-1}t(1-t)^{-1}$. This function g will be referred to throughout the remainder of Section 2.

Corollary 2.2.0.1 *For any fixed $\theta \in \mathbb{R}^+$, the d.f. for $\text{HM}_{k,n}(\theta)$ is*

$$F_{\text{HM}_{k,n}(\theta)}(t) = P(\text{HM}_{k,n}(\theta) \leq t) = F^{k*} \left(\frac{\theta kt}{\theta t + 1} \right), \quad (2.6)$$

where F^{k*} is the k -fold convolution of F_Y given in Lemma 1.

Proof The inverse of g is $g^{-1}(t) = t\theta(t\theta + 1)^{-1}$ and

$$F_{\text{HM}_{k,n}(\theta)}(t) = P(g(\bar{Y}_n) \leq t) = P(\bar{Y}_n \leq g^{-1}(t)) = F^{n*} \left(\frac{\theta nt}{\theta t + 1} \right). \quad \square$$

2.2.1 Efficiency of the Harmonic Moment Estimator

Efficiency properties of the proposed harmonic moment estimator, $\text{HM}_{k,n}(\theta)$, given in (3.9), are described in the section. For notational convenience, define $\mu := E(\bar{Y}_k) = \frac{\alpha\theta}{\alpha\theta+1}$,

and $\frac{\tau^2}{k} := \text{Var}(\bar{Y}_k) = \frac{\alpha\theta}{k(\alpha\theta+1)^2(\alpha\theta+2)} < \infty$. Consistency of and asymptotic normality of $\text{HM}_{k,n}(\theta)$ are shown first.

Theorem 2.2.1.1 *For any fixed $\theta \in \mathbb{R}^+$,*

$$\text{HM}_{k,n}(\theta) \xrightarrow{\text{P}} \alpha. \quad (2.7)$$

Proof From the weak law of large numbers $\bar{Y}_k \xrightarrow{\text{P}} \mu$. The continuity of g then gives

$$\text{HM}_{k,n}(\theta) \stackrel{\text{d}}{=} g(\bar{Y}_k) \xrightarrow{\text{P}} g(\mu) = \theta^{-1}\mu(1-\mu)^{-1} = \alpha. \quad \square$$

Theorem 2.2.1.2 *For any fixed $\theta \in \mathbb{R}^+$,*

$$\sqrt{k}(\text{HM}_{k,n}(\theta) - \alpha) \xrightarrow{\text{d}} \text{N}\left(0, \frac{\alpha(\alpha\theta+1)^2}{\theta(\alpha\theta+2)}\right). \quad (2.8)$$

Proof From the CLT we have $\sqrt{k}(\bar{Y}_k - \mu) \xrightarrow{\text{d}} \text{N}(0, \tau^2)$. An application of the Mann-Wald theorem yields

$$\sqrt{k}(\text{HM}_{k,n}(\theta) - \alpha) \stackrel{\text{d}}{=} \sqrt{k}(g(\bar{Y}_k) - g(\mu)) \xrightarrow{\text{d}} \text{N}\left(0, \{\tau g'(\mu)\}^2\right).$$

Note that $\{g'(t)\}^2 = \theta^{-1}(1-t)^{-2}$ so that $\{g'(\mu)\}^2 = \theta^{-2}(\alpha\theta+1)^4$. Hence

$$\tau^2\{g'(\mu)\}^2 = \left(\frac{\alpha\theta}{(\alpha\theta+1)^2(\alpha\theta+2)}\right) \frac{(\alpha\theta+1)^4}{\theta^2} = \frac{\alpha(\alpha\theta+1)^2}{\theta(\alpha\theta+2)}. \quad \square$$

From above we see that the asymptotic variance of $\text{HM}_{k,n}(\theta)$ achieves the Cramer-Rao lower bound as $\theta \uparrow \infty$. Also, from Theorem 2.2.1.2,

$$\text{HM}_{k,n}(\theta) \pm \Phi^{-1}(1-\beta) \sqrt{\frac{\text{HM}_{k,n}(\theta)(\text{HM}_{k,n}(\theta)\theta+1)^2}{k\theta(\text{HM}_{k,n}(\theta)\theta+2)}} \quad (2.9)$$

is a level β asymptotic confidence intervals for α , where Φ is the distribution function of a standard normal random variable.

Corollary 2.2.1.1 *For any $\varepsilon \in (0, 1)$, asymptotic relative efficiency of $\text{Hill}_{k,n}$ to $\text{HM}_{k,n}(\theta)$ is within ε of 1 by taking $\theta > (\sqrt{1/\varepsilon} - 1)/\alpha$.*

Proof From Theorem 2.2.1.2, the asymptotic relative efficiency of $\text{Hill}_{k,n}$ to $\text{HM}_{k,n}(\theta)$ is

$$\lim_{k(n) \rightarrow \infty} \frac{\text{AVar}(\text{Hill}_{k,n})}{\text{AVar}(\text{HM}_{k,n}(\theta))} = \frac{\theta\alpha(\alpha\theta + 2)}{(\alpha\theta + 1)^2} = 1 - (\alpha\theta + 1)^{-2}. \quad (2.10)$$

Then $(\alpha\theta + 1)^{-2} < \varepsilon \Leftrightarrow \theta > (\sqrt{1/\varepsilon} - 1)/\alpha$. \square

This corollary shows that the proposed estimator works very well in terms of efficiency for large θ . However, as we will see in the next section, there is then a price to pay in terms of robustness.

2.2.2 Robustness of the Harmonic Moment Estimator

Recall that while the asymptotic variance of the Hill estimator is equal to the CRLB, the Hill estimator has asymptotic contamination BP of 0, and hence is not robust in this sense. The following robustness result shows that the asymptotic contamination BP of $\text{HM}_{k,n}(\theta)$ is nonzero.

Theorem 2.2.2.1 *The estimator $\text{HM}_{k,n}(\theta)$ has asymptotic contamination BP of*

$$\varepsilon^* = \frac{\min\{\alpha\theta, 1\}}{\alpha\theta + 1} \in (0, 1/2] \quad (2.11)$$

for any fixed $\theta \in \mathbb{R}^+$. Moreover, ε^* achieves a maximum of $1/2$ when $\theta = \alpha^{-1}$.

Before proving Theorem 2.2.2.1, it is worth mentioning that the theorem says one should choose a value of θ to minimize ε^* , that is a function of the unknown tail index α . Of course this is impossible if α is unknown. One can however initially start with some reasonable estimator, say $\hat{\alpha}_0^* := \text{HM}_{k,n}(1)$, and calculate $\theta_1^* = 1/\hat{\alpha}_0^*$. Use θ_1^* to determine $\text{HM}_{k,n}(\theta_1^*)$. Call this $\hat{\alpha}_1^*$ and repeat the process. In the simulation study given in Section 3, this iterative procedure is used resulting in the following robust estimator for α :

$$\hat{\alpha} = \text{HM}_{k,n}(\theta^*) \quad \text{where} \quad \theta^* := \lim_m \theta_m^*. \quad (2.12)$$

Proof The approach described by Maronna et al (2006) is used here. Defining $\psi(y; \mu) := y - \mu$, we see from (3.9) that $\text{HM}_{k,n}(\theta)$ is the M -estimator corresponding to the estimating equation $\sum_1^k \psi(Y_i; \mu) = 0$ (note that $Y_i = (X^{(k+1)}/X^{(i)})^{1/\theta}$). From Maronna et al (2006), the asymptotic contamination BP of $\text{HM}_{k,n}(\theta)$ is

$$\varepsilon^* = \frac{\min\{-\psi(0; \mu), \psi(1; \mu)\}}{\psi(1; \mu) - \psi(0; \mu)} = \min\{\mu, 1 - \mu\} = \frac{\min\{\alpha\theta, 1\}}{\alpha\theta + 1} \in (0, 1/2].$$

It is easy to check that ε^* attains a maximum of $1/2$ when $\theta = \alpha^{-1}$. □

From above we see that the asymptotic contamination BP of $\text{HM}_{k,n}(\theta)$ approaches 0 as θ increases. The following shows that the Hill estimator is a special case of the harmonic moment estimator if we allow $\theta \in \mathbb{R}^+ \cup \{\infty\}$.

Corollary 2.2.2.1 *As $\theta \uparrow \infty$, the estimator $\text{HM}_{k,n}(\theta)$ becomes the Hill estimator. That*

is,

$$\lim_{\theta \uparrow \infty} \text{HM}_{k,n}(\theta) = \text{Hill}_{k,n}. \quad (2.13)$$

Proof Let $\theta \uparrow \infty$ in (3.9). □

The GES is another measure of robustness of an estimator and is a measure of the maximal impact on an estimator that results from the change in a single observation. The GES for the proposed estimator is given next.

Theorem 2.2.2.2 *The GES of $\text{HM}_{k,n}(\theta)$ is $(\alpha\theta + 1) \max\{\alpha, \theta^{-1}\}$.*

Proof Let ψ be as in Theorem 2.2.2.1. From Maronna et al (2006), the GES then is

$$\sup_{y \in [0,1]} \left| \frac{\psi(y; \mu)}{\int_0^1 \frac{\partial}{\partial \alpha} \psi(y; \mu) dF_Y(y)} \right| = \frac{(\alpha\theta + 1)^2}{\theta} \sup_{y \in [0,1]} \left| y - \frac{\alpha\theta}{\alpha\theta + 1} \right| = (\alpha\theta + 1) \max\{\alpha, \theta^{-1}\}.$$

□

Taking the results of this section together suggest that θ should not be chosen too large if robustness is a concern. Recall from the previous section, however that the efficiency of $\text{HM}_{k,n}(\theta)$ increases as a function of θ . The question of choosing θ is addressed in the following section. It should be noted here that robust estimation of the Pareto tail index parameter is also discussed in Brazauskas and Serfling (2000), Peng and Welsh (2001), Juarez and Schucany (2004), and Finkelstein et al (2006).

2.2.3 Investigating the Choice of θ

In the previous two subsections we have seen how a choice of the parameter θ allows one to control the tradeoff between efficiency and robustness. In this section, we investigate how θ can be chosen to obtain specified properties. The bias of $\text{HM}_{k,n}(\theta)$ is investigated first. An application of Jensen's inequality tells us that $E[\text{HM}_{k,n}(\theta)] \geq \alpha$ for any fixed $\theta \in \mathbb{R}^+$. The following lemma gives an approximation for the bias of $\text{HM}_{k,n}(\theta)$.

Lemma 2.2.3.1 *For any fixed $\theta \in \mathbb{R}^+$,*

$$E[\text{HM}_{k,n}(\theta)] = \alpha + \frac{\alpha(\alpha\theta + 1)}{k(\alpha\theta + 2)} + O\left(\frac{1}{k^2}\right). \quad (2.14)$$

Proof Expanding $\text{HM}_{k,n}(\theta) \stackrel{d}{=} g(\bar{Y}_k)$ in a Taylor series, we have

$$g(\bar{Y}_k) = g(\mu) + g'(\mu)(\bar{Y}_k - \mu) + \frac{1}{2}g''(\mu)(\bar{Y}_k - \mu)^2 + O_p\left(\frac{1}{k\sqrt{k}}\right), \quad (2.15)$$

where we say U_n is $O_p(n^\beta)$ if for each $\varepsilon > 0$ there exists a real number $\delta = \delta_\varepsilon > 0$ and a natural number $m = m_\varepsilon$ such that for every $n \geq m$

$$P\left\{|U_n/n^\beta| < \delta\right\} > 1 - \varepsilon.$$

Taking expected values in (2.15) gives

$$E[\text{HM}_{k,n}(\theta)] = E[g(\bar{Y}_k)] = g(\mu) + \frac{1}{2}g''(\mu)\frac{\tau^2}{k} + O\left(\frac{1}{k^2}\right).$$

Noting that $g''(t) = 2\theta^{-1}(1-t)^{-3}$, we have

$$\begin{aligned} E[\text{HM}_{k,n}(\theta)] &= \alpha + \frac{1}{2} \left(\frac{2(\alpha\theta + 1)^3}{\theta} \right) \left(\frac{\alpha\theta}{k(\alpha\theta + 1)^2(\alpha\theta + 2)} \right) + O\left(\frac{1}{k^2}\right) \\ &= \alpha + \frac{\alpha(\alpha\theta + 1)}{k(\alpha\theta + 2)} + O\left(\frac{1}{k^2}\right). \quad \square \end{aligned}$$

From the above lemma and the asymptotic variance of $\text{HM}_{k,n}(\theta)$ given in (2.8), the following theorem provides a choice of the parameter θ that makes $\text{HM}_{k,n}(\theta)$ optimal in a mean squared error sense.

Theorem 2.2.3.1 *A choice of θ given by*

$$\theta = \frac{\sqrt{k^2 + 8k} + k}{2\alpha} < \infty \quad (2.16)$$

minimizes

$$\left(\frac{\alpha(\alpha\theta + 1)}{k(\alpha\theta + 2)} \right)^2 + \frac{\alpha(\alpha\theta + 1)^2}{k\theta(\alpha\theta + 2)} \approx [\text{Bias}(\text{HM}_{k,n}(\theta))]^2 + \text{Var}(\text{HM}_{k,n}(\theta)) = \text{MSE}(\text{HM}_{k,n}(\theta)). \quad (2.17)$$

Before proving Theorem 2.2.3.1, notice that θ^\dagger depends on the true unknown α , and therefore can never be determined exactly. The same procedure that was used to obtain $\text{HM}_{k,n}(\theta^*)$ in (2.12) can be used here to obtain a MSE optimal version of $\text{HM}_{k,n}(\theta)$. That is, start with some reasonable estimator of α , say $\hat{\alpha}_0^\dagger := \text{HM}_{k,n}(1)$, and calculate $\theta_1^\dagger = 1/\hat{\alpha}_0^\dagger$. Use θ_1^\dagger to determine $\text{HM}_{k,n}(\theta_1^\dagger)$. Call this $\hat{\alpha}_1^\dagger$ and repeat the process. In the simulation study given in Section 3, this iterative procedure is used resulting in the following estimator for α :

$$\hat{\alpha} = \text{HM}_{k,n}(\theta^\dagger) \quad \text{where} \quad \theta^\dagger := \lim_m \theta_m^\dagger. \quad (2.18)$$

Proof The approximation for the variance and bias of $\text{HM}_{k,n}(\theta)$ is from (2.8) and (2.14).

Calculus shows that θ from (2.16) minimizes the MSE in (2.17). \square

It is worth noting that although θ^\dagger makes $\text{HM}_{k,n}(\theta)$ optimal in a mean squared error sense, θ^\dagger is $O(k)$. Hence if robustness is a concern, the results from Section 2.2 can be incorporated into the choice of θ . Corollary 2.2.1.1 can also be considered if one is looking for a specific level of efficiency.

If desired, one can also choose to average over a range of θ values. That is, estimating α by

$$\overline{\text{HM}_{k,n}(\theta)} := \int_B \text{HM}_{k,n}(\theta) f(\theta) d\theta, \quad (2.19)$$

where f is a density function with support $B \subset \mathcal{B}(\mathbb{R}^+)$. In light of the bias term in (2.14), a bias adjusted estimator given by

$$\widetilde{\text{HM}_{k,n}}(\theta) := \text{HM}_{k,n}(\theta) \left(1 - \frac{\theta \text{HM}_{k,n}(\theta) + 1}{k[\theta \text{HM}_{k,n}(\theta) + 2]} \right) \quad (2.20)$$

can also be used. Using the delta method, the variance of $\widetilde{\text{HM}_{k,n}}(\theta)$ is approximately $(f'[E\bar{Y}_k])^2 \text{Var}(\bar{Y}_k) = (f'(\mu))^2 \tau^2 / k$ where $f(t) := g(t) \left(1 - \frac{\theta g(t) + 1}{k(\theta g(t) + 2)} \right)$. This result can be used to construct confidence intervals similar to Equation (2.9).

2.3. Simulation Study

2.3.1 Description

In this section $\text{HM}_{k,n}(\theta)$ given in (3.9) is compared to some other well known estimators. Besides the Hill estimator already mentioned (see (2.3)), the following tail index estimators are also included in the simulation study: method of moments (MOM), method

of probability weighted moment (PWM), and DEdH proposed in Dekkers and de Haan (1989). The simulation was done in R, and the function `fitgpd` function found in the POT package (see Ribatet (2006)) was used to calculate the MOM and PWM estimators. Two different distributions were simulated from, a Pareto distribution with tail index $\alpha = 2.5$:

$$F(x) = 1 - x^{-2.5}, \quad x \geq 1, \quad (2.21)$$

and a Burr distribution with tail index $\alpha = 2$:

$$F(x) = 1 - \left(1 + \frac{\sqrt{x}}{2}\right)^{-4}. \quad (2.22)$$

The MOM, PWM, DEdH, and Hill estimators were chosen because, like $\text{HM}_{k,n}(\theta)$, they are also moment estimators in some sense. Although there are many natural ways of arriving at the Hill estimator, it can be obtained in the same way as $\text{HM}_{k,n}(\theta)$ by using the geometric mean instead of the harmonic mean. See Embrechts et al (1997), pages 330–336, for three other ways of arriving at the Hill estimator. The number k of upper order statistics was chosen to be small (relative to the sample size n) purposefully, since one working in extreme value theory often finds themselves working with few observations.

2.3.2 Results

Table 2.1 summarizes how the $\text{HM}_{k,n}(\theta)$ compares to some other tail index estimators in terms of bias, standard deviation (SD), and root MSE, using the distributions in (2.21) and (2.22). As we can see in Table 2.1, the Hill estimator and harmonic moment (HM) estimator outperform (with regards to bias, SD, and MSE) the MOM, DEdH, and PWM estimators.

Comparison of $HM_{k,n}(\theta)$ with other estimators: Bias, SD, and MSE

| number k of upper order statistics | | | | | number k of upper order statistics | | | | |
|---|-------|-------|-------|-------|---|-------|-------|-------|-------|
| Estimator | 5 | 10 | 15 | 20 | Estimator | 10 | 20 | 50 | 100 |
| MOM | 0.149 | 0.086 | 0.150 | 0.014 | MOM | 0.002 | 0.018 | 0.013 | 0.010 |
| DEdH | 0.140 | 0.011 | 0.062 | 0.446 | DEdH | 0.369 | 0.001 | 0.037 | 0.078 |
| PWM | 0.241 | 0.004 | 0.043 | 0.012 | PWM | 0.010 | 0.072 | 0.042 | 0.022 |
| Hill | 1.000 | 1.000 | 1.000 | 1.000 | Hill | 1.000 | 1.000 | 1.000 | 1.000 |
| $HM(\theta^*)^a$ | 1.129 | 1.129 | 1.127 | 1.125 | $HM(\theta^*)$ | 1.118 | 1.114 | 1.150 | 1.597 |
| $HM(\theta^\dagger)^b$ | 1.114 | 1.071 | 1.052 | 1.041 | $HM(\theta^\dagger)$ | 1.069 | 1.040 | 1.019 | 1.016 |
| (a) | | | | | (d) | | | | |
| number k of upper order statistics | | | | | number k of upper order statistics | | | | |
| Estimator | 5 | 10 | 15 | 20 | Estimator | 10 | 20 | 50 | 100 |
| MOM | 0.002 | 0.000 | 0.001 | 0.000 | MOM | 0.000 | 0.000 | 0.001 | 0.008 |
| DEdH | 0.002 | 0.000 | 0.001 | 0.001 | DEdH | 0.002 | 0.000 | 0.002 | 0.008 |
| PWM | 0.003 | 0.000 | 0.001 | 0.000 | PWM | 0.000 | 0.001 | 0.002 | 0.008 |
| Hill | 1.000 | 1.000 | 1.000 | 1.000 | Hill | 1.000 | 1.000 | 1.000 | 1.000 |
| $HM(\theta^*)$ | 0.832 | 0.857 | 0.861 | 0.865 | $HM(\theta^*)$ | 0.860 | 0.863 | 0.866 | 0.868 |
| $HM(\theta^\dagger)$ | 1.018 | 1.005 | 1.002 | 1.001 | $HM(\theta^\dagger)$ | 1.005 | 1.001 | 1.000 | 1.000 |
| (b) | | | | | (e) | | | | |
| number k of upper order statistics | | | | | number k of upper order statistics | | | | |
| Estimator | 5 | 10 | 15 | 20 | Estimator | 10 | 20 | 50 | 100 |
| MOM | 0.002 | 0.000 | 0.001 | 0.000 | MOM | 0.000 | 0.000 | 0.001 | 0.008 |
| DEdH | 0.003 | 0.000 | 0.001 | 0.001 | DEdH | 0.002 | 0.000 | 0.002 | 0.008 |
| PWM | 0.003 | 0.000 | 0.002 | 0.000 | PWM | 0.000 | 0.001 | 0.002 | 0.008 |
| Hill | 1.000 | 1.000 | 1.000 | 1.000 | Hill | 1.000 | 1.000 | 1.000 | 1.000 |
| $HM(\theta^*)$ | 0.853 | 0.870 | 0.871 | 0.872 | $HM(\theta^*)$ | 0.874 | 0.871 | 0.870 | 0.870 |
| $HM(\theta^\dagger)$ | 1.027 | 1.009 | 1.005 | 1.003 | $HM(\theta^\dagger)$ | 1.009 | 1.003 | 1.001 | 1.000 |
| (c) | | | | | (e) | | | | |

^aRobust version of $HM_{k,n}(\theta)$. See Theorem 2.2.2.1 and the algorithm leading to (2.12)

^b θ is chosen to minimize the asymptotic MSE of $HM_{k,n}(\theta)$. See Theorem 2.2.3.1 and the algorithm leading to (2.18)

TABLE 2.1: In (a), (b), and (c), $n = 100$ Pareto r.v.'s with d.f. given in equation (2.21) were generated $m = 100,000$ times. In (a) ((b), and (c), respectively), the sample bias (SD in (b), and root MSE in (c), respectively) was calculated for each of the 6 estimators, for each shown value of k upper order statistics. The table shows the absolute value of the sample bias (SD in (b), and root MSE in (c), respectively) of the Hill estimator divided by the absolute value of the sample bias (SD in (b), and root MSE in (c), respectively) of each other estimator. In (d), (e), and (f), the same is done using a Burr distribution with d.f. given in equation (2.22). In (d), (e), and (f), $n = 1,000$ was used.

In Table 2.1(a), the simulated distribution is Pareto and the bias of the Hill and HM estimators can be compared. The bias of the Hill estimator was 12-13% higher than the bias of HM for $\theta = \theta^*$, and 11.4%, 7.1%, 5.2%, and 4.1% higher than the bias of HM for $\theta = \theta^\dagger$ and $k = 5, 10, 15$, and 20.

A more interesting result is seen in Table 2.1(d), where generated values are from the Burr distribution. If too many upper order statistics are included, tail index estimators become biased. In the case of an exact Pareto distribution, of course there is no such thing as too many upper order statistics. Table 2.1(d) illustrates robustness of HM; i.e., when $\theta = \theta^*$ the bias of the Hill estimator grows faster than the bias of the HM estimator.

The SD of the Hill and HM estimators are almost identical when simulating from the Pareto (Table 2.1(b)) and Burr (Table 2.1(e)) distributions for $\theta = \theta^\dagger$, at each value of k . For the more robust version of HM, when $\theta = \theta^*$, the SD of the Hill estimator was roughly 14% less than the SD of the HM estimator.

Because of the small bias of both the Hill and HM estimators, the root MSE of these estimators is roughly equal to the SD of the estimators. Therefore the results in Table 2.1(c) and (f) are roughly the same as those in Table 2.1(b) and (e).

2.4. Threshold Selection: A new Sum Plot

In order to estimate the tail index parameter above a threshold $X^{(k+1)}$, one needs to determine an appropriate value of k . There exist a variety of diagnostic plots that aid in threshold selection include the Sum, Zipf, Hill, and empirical mean-excess plots. Other threshold selection procedures can be found in Hall (1990), Dekkers and de Haan (1993), Dupuis (1999) and Hsieh (1999).

The Sum Plot proposed by de Sousa and Michailidis (2004) is made up of the points

$$\{(k, S_k) : k = 1, \dots, n-1\}, \quad (2.23)$$

where $S_k := \sum_{i=1}^k i \log \frac{X^{(i)}}{X^{(i+1)}}$, for $k \in \{1, \dots, n-1\}$. Then conditional on $X^{(k+1)} > u$ in Equation (2.2), S_k has a Gamma distribution with mean k/α and the above plot should be linear in a k -range of upper order statistics where the Pareto assumption holds. After identifying a region of upper order statistics where the plot is linear, one may use these order statistics to estimate the tail index parameter. Sousa and Michailidis (2004) provide an algorithm for choosing k from the Sum Plot.

A similar technique is proposed here. Define $R_k := \sum_{i=1}^k (X^{(k+1)}/X^{(i)})^{1/\theta}$, $k \in \{1, \dots, n-1\}$. Then conditional on $X^{(k+1)} > u$ in Equation (2.2), we have $E[R_k] = k \left(\frac{\alpha\theta}{\alpha\theta+1} \right)$. Then, as with Sum Plot given in (2.23), the plot

$$\{(k, R_k), k = 1, 2, \dots, n-1\} \quad (2.24)$$

should be linear in the appropriate region of upper order statistics.

Next, the two sum plots in Equations (2.23) and (2.24) are compared in a few examples using the following distribution:

$$F(x) = \begin{cases} 1 - e^{-\gamma x}, & \text{if } x \leq D, \\ 1 - cx^{-\alpha}, & \text{if } x \geq D. \end{cases} \quad (2.25)$$

This distribution was designed so that the threshold beyond which one should estimate

Comparison of Sum Plots

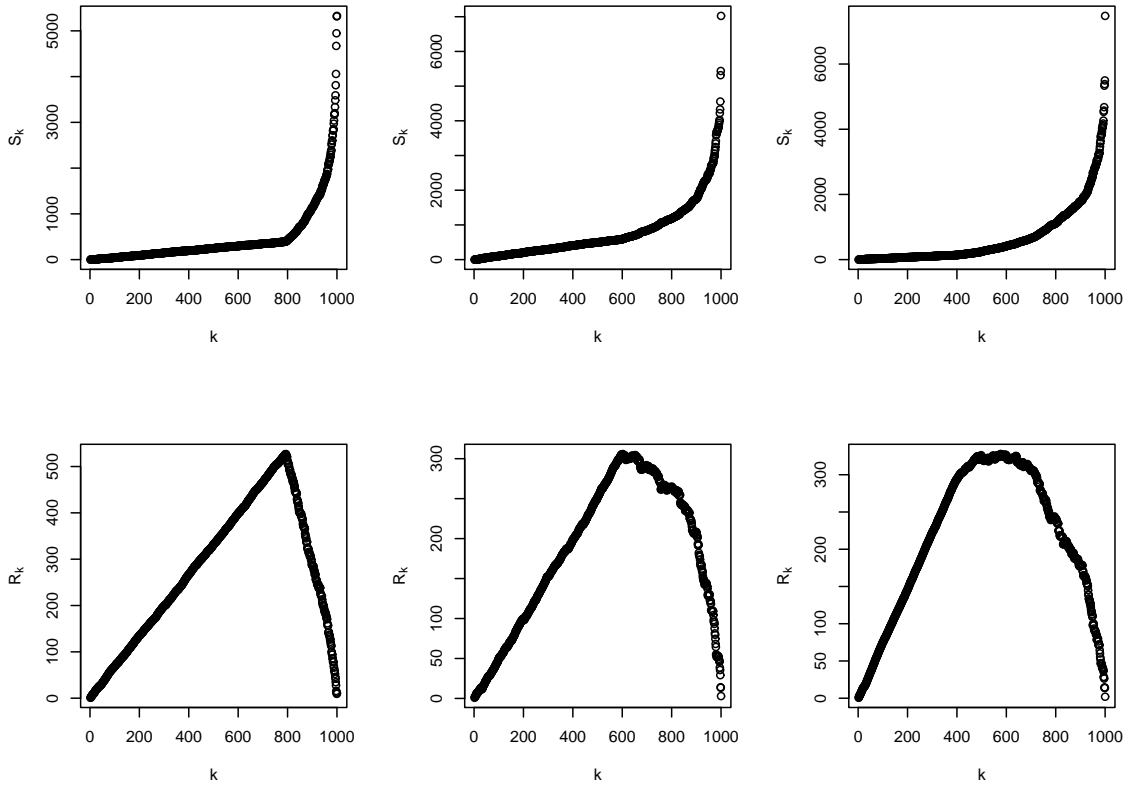


FIGURE 2.1: Let $\boldsymbol{\rho} := (\alpha, \gamma, 1 - F(D))$ in Equation (2.25). Above plots show $\{(k, S_k)\}_1^{n-1}$ and $\{(k, R_k)\}_1^{n-1}$ for $n = 1,000$ simulated values from F in Equation (2.25) with Left: $\boldsymbol{\rho} = (2, 1, 0.2)$; Center: $\boldsymbol{\rho} = (1, 3, 0.4)$; and Right: $\boldsymbol{\rho} = (3, 1/3, 0.6)$. A fixed value of $\theta \equiv 1$ was used for each R_k . The plots should break linearity around $k = 800$ (left), $k = 600$ (center), and $k = 400$ (right).

α is known, and equal to D . Note that F is exponential below D and Pareto above D .

For the plots to be a useful diagnostic tool, they should be linear in the region of upper

order statistics greater than D , and one should be able to easily estimate the number k of order statistics greater than D .

In Figure 2.1, D was chosen as the 0.8 (left), 0.6 (center), and 0.4 (right) quantiles of F . For example, for the simulated values from F in the plots on the left, we expect that 20% of the values were greater than D . We can see clear breaks from linearity in all three of the $\{(k, R_k)\}_1^{n-1}$ plots. Furthermore, $\{(k, R_k)\}_1^{n-1}$ seems to be at least as useful as $\{(k, S_k)\}_1^{n-1}$ for helping to choose k .

2.5. Applications in Insurance

Extreme value theory is often used when modeling large insurance losses. See, for example, Beirlant and Teugels (1992), McNeil (1997), Embrechts et al (1999), Beirlant et al (2001), and Cebrián et al (2003). In particular, loss distributions with heavy tails are common, and so estimation of the tail index parameter becomes an important problem. In this section the Secura Belgian Re automobile data, looked at in a case study in Beirlant et al (2004), is used to illustrate the proposed harmonic moment tail index estimator. See Beirlant et al (2004) for a more detailed analysis.

The 371 automobile claims, from 1988 until 2001, in this data set are all greater than 1,200,000 Euro, and have been adjusted for inflation and other things. In Beirlant et al (2004), estimation of the net premium of an excess-of-loss, above a retention level R ,

reinsurance contract is of interest. This net premium is denoted $\Pi(R)$. Letting X be the claim size r.v., the net premium is given by

$$\Pi(R) = E \{ (X - R) | X > R \} P(X > R)$$

and can be estimated by

$$\hat{\Pi}(R) = \frac{R}{\hat{\alpha}_k - 1} \left(\frac{R}{X^{(k+1)}} \right)^{-\hat{\alpha}_k} \frac{k}{n}, \quad \text{for } R \geq X^{(k+1)}. \quad (2.26)$$

In Figure 2.2 several diagnostic plots are shown for the Secura Belgian Re automobile

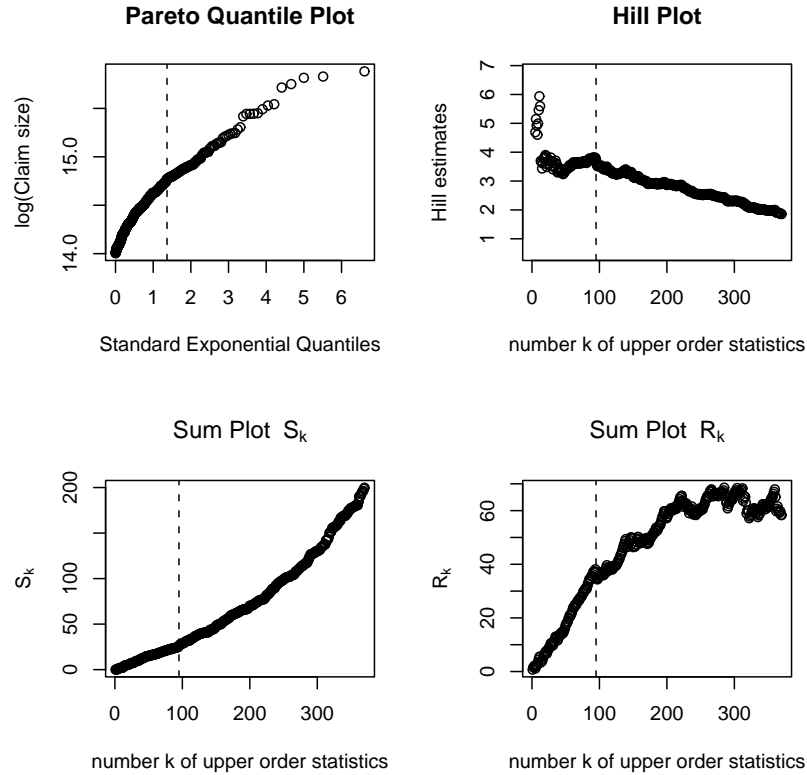


FIGURE 2.2: Diagnostic plots for the Secura Belgian Re automobile claims data. The vertical line shows the number of upper order statistics chosen ($k = 95$) by in Beirlant et al (2004).

data that aid in choosing the number of upper order statistics to use. The Pareto quantile

| Estimates for $\Pi(R)$ | | | |
|--|---------------------|--------------|---------------------------|
| R | $\theta = \theta^*$ | $\theta = 1$ | $\theta = \theta^\dagger$ |
| 3,000,000 | 154,727.7 | 162,699.6 | 163,812.0 |
| 3,500,000 | 100,498.8 | 107,279.7 | 108,230.8 |
| 4,000,000 | 69,154.6 | 74,789.7 | 75,584.4 |
| 4,500,000 | 49,731.1 | 54,405.6 | 55,068.3 |
| 5,000,000 | 37,028.6 | 40,928.1 | 41,483.7 |
| 7,500,000 | 11,901.4 | 13,686.1 | 13,945.5 |
| 10,000,000 | 5,319.2 | 6,291.2 | 6,434.6 |

TABLE 2.2: Estimates of net premium, $\Pi(R)$, in Equation (2.26) using $\text{HM}_{k,n}(\theta)$ to estimate α with different choices of θ . See (2.12), and (2.18), for definitions of θ^* and θ^\dagger . $\theta = 1$ corresponds to the standard harmonic moment estimator.

plot (upper left) should be ultimately linear for distributions satisfying Equation (2.1). In the Hill plot (upper right), we look for the largest k where the plot appears stable. The sum plots (bottom) were discussed in Section 4. The dotted vertical lines correspond to the k value of 95 that was chosen in Beirlant et al (2004), and is used here also.

Using $k = 95$, different versions of $\text{HM}_{k,n}(\theta)$ were determined and used for $\hat{\alpha}_k$ in Equation (2.26). The results are given in Table 2.2, and are similar to those given in Beirlant et al (2004).

In Figure 2.3 estimated conditional tail probabilities are plotted along with empirical probabilities. The conditional tail probabilities estimates are determined using

$$1 - \hat{F}_u(x) \equiv \hat{P}(X > x \mid X > u) = \left(\frac{x}{u}\right)^{-\hat{\alpha}}, \text{ for } x > u, \quad (2.27)$$

where $k = 95$, $\theta = 1$, $u = x^{(k+1)} = 2,580,026$, $\hat{\alpha} = \text{HM}_{k,n}(\theta) = 3.7$.

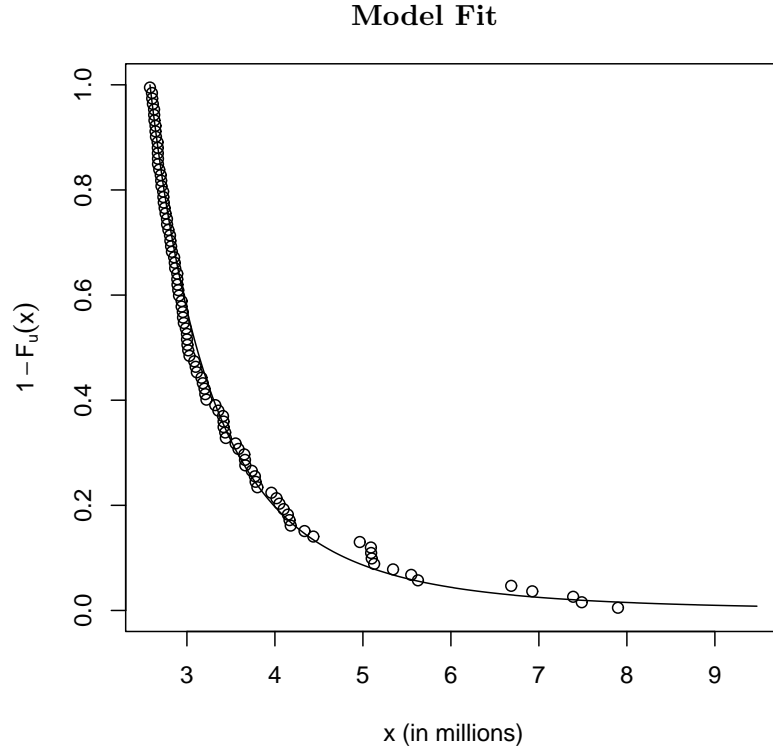


FIGURE 2.3: Pareto tail fit to the loss data using $k = 95$ and $\theta = 1$ in Equation (2.27). The model fit is given by the solid line and the points are the empirical tail probabilities.

2.6. Conclusion

In this paper, a new family of tail index estimators given by

$$\{\text{HM}_{k,n}(\theta) : k, n \text{ integer}; 1 \leq k \leq n - 1, \theta \in \mathbb{R}^+ \cup \{\infty\}\}$$

has been discussed. This set contains the popular Hill estimator, a MLE estimator for α ; an estimator with the easy interpretation as being obtained by matching theoretical and empirical harmonic means ($\theta = 1$), as well as more robust estimators of α (for example, when $\theta = \theta^*$). The flexibility of the proposed estimator allows one to compare many different estimates for α just by varying θ .

The proposed estimator seems to work particularly well when k is small, and the robustness properties should make it useful for distributions F satisfying Equation (2.1) when ℓ is not constant. These features are illustrated in the simulation study when k is small, and when $\ell(x) = (1 + 0.5x^{-0.5})^{-4}$ for the Burr distribution.

The new diagnostic sum plot given in (2.24) is based on the distributional properties of the Y_i 's in Lemma 2.2.0.1. While there are many diagnostic plots that can, and should, be studied when choosing the number of upper order statistics to use in tail index estimation, many existing plots can be unstable or hard to interpret. The new sum plot provides an additional reliable tool to aid in threshold selection.

3. EXTREME VALUE ANALYSIS FOR PARTITIONED INSURANCE LOSSES

John B. Henry III and Ping-Hung Hsieh

Abstract

The heavy-tailed nature of insurance claims requires that special attention be put into the analysis of the tail behavior of a loss distribution. It has been demonstrated that the distribution of large claims of several lines of insurance have Pareto-type tails. As a result, estimating the tail index, which is a measure of the heavy-tailedness of a distribution, has received a great deal of attention. Although numerous tail index estimators have been proposed in the literature, many of them require detailed knowledge of individual losses and are thus inappropriate for insurance data in partitioned form. In this study we bridge this gap by developing a tail index estimator suitable for partitioned loss data. This estimator is robust in the sense that no particular global density is assumed for the loss distribution. Instead we focus only on fitting the model in the tail of the distribution where it is believed that the Pareto-type form holds. Strengths and weaknesses of the proposed estimator are explored through simulation and an application of the estimator to real world partitioned insurance data is given.

3.1. Introduction

The heavy-tailed nature of insurance claims requires that special attention be put into the analysis of the tail of a loss distribution. Since a few large claims can significantly impact an insurance portfolio, statistical methods that deal with extreme losses have become necessary for actuaries. For example, in order to price certain reinsurance treaties, it is often necessary for actuaries to model losses in excess of some high threshold value, i.e., to model the largest k upper order statistics. Beirlant and Teugels (1992), McNeil (1997), Embrechts et al (1999), Beirlant et al (2001), Cebrián et al (2003) and Beirlant et al (2004) provide additional examples where statistical methods were developed to deal with extreme insurance losses.

Extreme value theory has become one of the main theories in developing statistical models for extreme insurance losses. The theory states that the tail of a typical loss distribution $F_X(x)$ can be approximated by a Pareto function. That is, $1 - F_X(x) \approx x^{-\alpha}$, where the parameter α is known in the literature as the Pareto tail index that measures the heavy-tailedness of the loss distribution. See, for example, Finkelstein et al (2006). Many distributions commonly seen in modeling insurance losses have Pareto type tails. They include the Pareto, generalized Pareto, Burr, Fréchet, half T, F, inverse gamma, and log gamma distributions. Following the theory, an actuary may assume that the tail of the loss distribution, where extreme losses occur, can be approximated by a Pareto function without making specific assumption on the global density. With an estimate of the Pareto

index parameter, the actuary can then estimate quantities of interest that are related to extreme losses, e.g., expected loss above a high retention limit. The approximation of Pareto function has been demonstrated to be reasonable for many lines of insurance. Numerous tail index estimators have also been proposed in the literature including earlier contributions by Hill (1975) and Pickands (1975) in which the Hill estimator has become somewhat of a benchmark to which later proposed estimators are compared. A survey of existing estimators including their advantages and disadvantages can be found in Brazauskas and Serfling (2000), Hsieh (2002), and Beirlant et al (2004).

Insurance loss data reported in partitioned form are common in practice. The frequencies of losses occurred in certain loss intervals for numerous lines of insurance can often be found in companys' reports or in government publications. Individual loss data are typically proprietary to the company and may not be available to its competitors in the industry. Despite the number of tail index estimator proposed in the literature, many, if not all, of them require the use of individual loss data, and thus, are inappropriate for tail index estimation under the constraint of partitioned data. This article intends to expand the horizon of tail index estimation by applying extreme value theory to partitioned loss data. The main objective is to propose a robust tail index estimator for partitioned loss data. The estimator is robust in the sense that no global density is assumed and the Pareto function is used to approximate the tail of a large class of distributions commonly used in modeling insurance loss data. This approach is advantageous because fitting a global density to losses can lead to errors when making tail inference in the event that the

true loss distribution does not have the assumed density. Instead, we rely on the extreme value theory and focus only on fitting the tail of the distribution without assuming a specific global density. In addition, we will demonstrate the loss of efficiency by using the partitioned data versus individual data through simulation.

The remainder of the paper is arranged as follows. Several tail index estimators are reviewed in Section 3.2.. Except for the Hill and Pickands estimators, both of which have historical values and the former is served as a benchmark in our simulation, the rest of the review intends to be a supplement to the excellent review of Brazauskas and Serfling (2000), Hsieh (2002), and Beirlant et al (2004). The derivation of the proposed estimator and an examination of its theoretical properties are worked out in Section 3.3.. In Section 3.4., a simulation study is conducted to assess the performance of the proposed estimator. Two questions that guide the design of the simulation are (1) what is the efficiency lost by using data in partitioned form? and (2) what is the penalty of model misspecification? The simulation results are discussed in Section 3.5.. Insurance applications are given in Section 3.6. using actual grouped insurance losses, followed by concluding remarks in Section 3.7..

3.2. Literature Review

In this section we consider tail index estimators for a loss random variable (r.v.) X taking values on the positive real line \mathbb{R}^+ with nondegenerate distribution function F_X . We

assume that the loss distribution has a Pareto-like tail in the sense that

$$P(X > x) \sim Cx^{-\alpha}, \text{ as } x \rightarrow \infty, \quad (3.1)$$

where $\alpha > 0$ and \sim indicates the limit of the ratio of the two functions is equal to one. In this case the probability that a loss exceeds a level x can be closely approximated by $Cx^{-\alpha}$ when x is larger than some threshold D . Survival functions for random variables satisfying (3.1) are often written as $1 - F_X(x) = \ell(x)x^{-\alpha}$ where $\alpha > 0$ is the (upper) tail index, and $\ell(x)$ is a function slowly varying at infinity. Note that a Lebesgue measurable function $\ell : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ is slowly varying at infinity if $\lim_{x \rightarrow \infty} \frac{\ell(tx)}{\ell(x)} = 1$ for all $t > 0$. We will denote the tail probability function by $\bar{F}_X(x) := 1 - F_X(x)$. Let $\{X_k : 1 \leq k \leq n\}$ be a sequence of independent copies of X and denote the descending order statistics by $X^{(1)} \geq X^{(2)} \geq \dots \geq X^{(n)}$.

In the following subsections, we discuss several estimators for the tail index α . Some noteworthy estimators that are not discussed below are the method of moments, probability-weighted moments, elemental percentile, Bayes estimator with conjugate priors and hybrid estimators. A description of these can be found, for example, in Hsieh (2002) and the references therein.

3.2.1 The Hill and Pickands Estimators

Hill (1975) proposed the tail index estimator

$$\hat{\alpha}_H = \frac{k+1}{\sum_{i=1}^k i \log \left(\frac{X^{(i)}}{X^{(i+1)}} \right)} \quad (3.2)$$

based on a maximum likelihood argument where $k \in \{1, 2, \dots, n-1\}$. The Hill estimator is closely related to the mean excess function $e(u) = E\{X - u \mid X > u\}$. In particular, the *empirical* mean excess function is given by $e_n(u) = [\text{card}\Lambda_n(u)]^{-1} \sum_{j \in \Lambda_n(u)} (X_j - u)$, where $\Lambda_n(u) = \{j : X_j - u > 0, j = 1, \dots, n\}$. Then, letting $e_n^*(u)$ denote the empirical mean excess function of the log transformed variables, we have $e_n^*(\log X^{(k+1)}) = \frac{1}{k} \sum_{i=1}^k (\log X^{(i)} - \log X^{(k+1)})$. As a result, we see that $\hat{\alpha}_H = \frac{k+1}{k} e_n^*(\log X^{(k+1)})^{-1}$. That is, the Hill estimator is asymptotically equal to the reciprocal of the empirical mean excess function of $\log X$ evaluated at the threshold $\log X^{(k+1)}$.

An important feature of the Hill estimator to keep in mind is the variance-bias tradeoff that occurs when choosing the number of upper order statistics to use. Choosing too many of the largest order statistics can lead to a biased estimator while too few increases the variability of the estimator. See Embrechts et al (1997) for a further variance-bias tradeoff discussion and Hall (1990), Dekkers and de Haan (1993), Dupuis (1999) and Hsieh (1999) for methods for determining the number of upper order statistics or threshold to use. Properties of the Hill estimator can be found in Embrechts et al (1997) and the references therein.

Pickands (1975) proposed an estimator that matches the 0.5 and 0.75 quantiles of the generalized Pareto distribution with quantile estimates. More specifically, for a GPD r.v.

X with distribution function

$$G(x; \xi, \sigma) = 1 - \left(1 + \frac{\xi x}{\sigma}\right)^{-1/\xi} 1_{(0, \infty)}(x),$$

it is easy to show that

$$\frac{G^{-1}(0.75) - G^{-1}(0.5)}{G^{-1}(0.5)} = 2^\xi.$$

Then denoting 0.5 and 0.75 quantile estimates by \hat{q}_1 and \hat{q}_2 , respectively, we have

$$\hat{\xi} = \frac{\log \left(\frac{\hat{q}_2 - \hat{q}_1}{\hat{q}_1} \right)}{\log 2}.$$

Pickands proposed, for n independent copies of X , using $\hat{q}_1 = X^{(m)} - X^{(4m)}$ and $\hat{q}_2 = X^{(2m)} - X^{(4m)}$ where $n \gg m \geq 1$. Then noting that the tail index for a GPD r.v. is given by $\alpha = 1/\xi$, the resulting tail index estimate is, for $X^{(k)} \geq D$,

$$\hat{\alpha}_P = \frac{\log 2}{\log \left(\frac{X^{(m)} - X^{(2m)}}{X^{(2m)} - X^{(4m)}} \right)}, \quad (3.3)$$

where $k \geq 4m \geq 4$.

For consistency and asymptotic results, see Dekkers and de Haan (1989). While the simplicity of the Pickands estimator is an attractive feature, it makes use of only 3 upper order statistics and can have a large asymptotic variance. Generalized versions of the Pickands estimator can be found, for example, in Seagers (2005). See Section 3.2.2.

3.2.2 Some Recent Tail Index Estimators

3.2.2.1 Censored Data Estimator

In the case of moderate right censoring, Beirlant and Guillou (2001) proposed an estimator based on the slope of the Pareto quantile plot, excluding the censored data. This can be

useful in situations when there has been a policy limit or when a reinsurer has covered losses in the portfolio exceeding some well defined retention level. Letting N_c denote the number of censored losses, the estimator is

$$\hat{\alpha}_{N_c}(k) = \frac{k - N_c}{\sum_{i=N_c+1}^k \log \frac{X^{(i)}}{X^{(k+1)}} + N_c \log \frac{X^{(N_c+1)}}{X^{(k+1)}}}, \quad (3.4)$$

where $k \in \{N_c+1, \dots, n-1\}$. This estimator is equivalent to the Hill estimator (except for the change from $k+1$ to k , which is asymptotically negligible) in the case of no censoring (i.e., $N_c = 0$). It is argued by Beirlant and Guillou (2001) that typically no more than 5% of observations should be censored for an effective use of this method.

3.2.2.2 Location Invariant Estimators

It is pointed out by Fraga Alves (2001) that, for modeling large claims in an insurance portfolio, it is desirable for an estimator of α to have the same distribution for the excesses taken over any possible fixed deductible. For this reason, location invariance is clearly a desirable property for an estimator of α . Fraga Alves (2001) introduced a Hill-type estimator that is made location invariant by a random shift. The location invariant estimator is

$$\hat{\alpha}_{k_0, k} = \frac{k_0}{\sum_{i=1}^{k_0} \log \frac{X^{(i)} - X^{(k+1)}}{X^{(k_0+1)} - X^{(k+1)}}}, \quad (3.5)$$

where k_0 is a secondary value chosen with $k_0 < k$. An algorithm is included in Fraga Alves (2001) to estimate the optimal k_0 , and to make a bias correction adjustment to $\hat{\alpha}_{k_0, k}$.

Generalized Pickands estimators described in Seagers (2005) are also location invariant and are linear combinations of log-spacings of order statistics. In particular, let Λ denote the collection of all signed Borel measures λ on $(0, 1]$ such that

$$\lambda((0, 1]) = 0, \quad \int \log(1/t)|\lambda|(dt) < \infty, \quad \text{and}, \quad \int \log(1/t)\lambda(dt) = 1.$$

Then for $\lambda \in \Lambda$ and $0 < c < 1$, the generalized Pickands estimators are given by

$$\hat{\alpha}_k(c, \lambda) = \left(\sum_{i=1}^k \left[\lambda\left(\frac{i}{k}\right) - \lambda\left(\frac{i-1}{k}\right) \right] \log \left(X^{(1+\lfloor cj \rfloor)} - X^{(i+1)} \right) \right)^{-1}. \quad (3.6)$$

See Seagers (2005) for examples using different measures λ and theoretical properties of the generalized Pickands estimators. See also Drees (1998) for a general theory of location and scale invariant tail index estimators that can be written as Hadamard differentiable continuous functionals of the empirical tail quantile function.

3.2.2.3 Generalized Median Estimator

A class of generalized median (GM) estimators were proposed by Brazauskas and Serfling (2000) with the goal of retaining a relatively high degree of efficiency while also being adequately robust. The GM estimator is found by considering, for $X^{(k)} \geq D$ and $r \in \{2, \dots, k\}$, the median of a kernel h evaluated over all $\binom{k}{r}$ subsets of $X^{(1)}, \dots, X^{(k)}$. The GM estimator is then given by

$$\hat{\alpha}_r = \text{med}\{h(X^{(i_1)}), \dots, h(X^{(i_r)})\}, \quad (3.7)$$

where $\{i_1, \dots, i_r\}$ corresponds to a set of distinct indices from $\{1, \dots, k\}$. Examples of kernels h , properties of the GM estimators, and comparison between the GM estimators and several other estimators can be found in Brazauskas and Serfling (2000).

3.2.2.4 Probability Integral Transform Statistic Estimator

Finkelstein et al (2006) describe a probability integral transform statistic (PITS) estimator for the tail-index parameter of a Pareto distribution. They develop the PITS estimator through an easily understandable and sound probabilistic argument. The PITS estimator is shown to be comparable to the best robust estimators. Consider first a random sample of Pareto random variables X_1, \dots, X_n , each with common distribution function $F(x) = 1 - (D/x)^\alpha$ for $x \geq D$ where $D > 0$ is known and $\alpha > 0$. Then defining

$$G_{n,t}(\beta) = \frac{1}{n} \sum_{i=1}^n \left(\frac{D}{X_i} \right)^{\beta t},$$

where $t > 0$, observe that

$$G_{n,t}(\alpha) = \frac{1}{n} \sum_{i=1}^n \bar{F}(X_i)^t \stackrel{d}{=} \frac{1}{n} \sum_{i=1}^n U_i^t,$$

where U_1, \dots, U_n are i.i.d Uniform(0,1) random variables. Applying the Strong Law of Large Numbers yields

$$G_{n,t}(\alpha) \xrightarrow{p} E(U_1^t) = (t+1)^{-1}.$$

Using the idea of method of moment estimation, the PITS estimator is the solution of the equation $G_{n,t}(\beta) = (t+1)^{-1}$. The tuning parameter $t > 0$ is used to adjust between robustness and efficiency. See Finkelstein et al (2006) for details. In the case D is unknown,

one can consider

$$G_{n,t,k}(\beta) := \frac{1}{n} \sum_{i=1}^k \left(\frac{X^{(k+1)}}{X^{(i)}} \right)^{\beta t},$$

for $k \in \{1, 2, \dots, n-1\}$ and use the same approach to arrive at a PITS estimator for the tail-index α .

3.3. Tail Index Estimator for Partitioned Data

Let $\{X_k : 1 \leq k \leq n\}$ be a sequence of independent copies of a loss random variable X satisfying (3.1). Suppose that losses are grouped into classes $\{I_i = (a_i, a_{i-1}]\}_{i=1, \dots, g}$, where $\infty = a_0 > a_1 > \dots > a_g > 0$. Assuming the loss distribution has the Pareto-type form above a threshold D , we take $0 < D \leq a_k$ without loss of generality for some $k \in \{2, 3, \dots, g\}$. We let N_1, \dots, N_g denote the frequencies with which (X_1, \dots, X_n) take values in $\{I_i = (a_i, a_{i-1}]\}_{i=1, \dots, g}$. That is, $N_i = \text{card}\{j : a_{i+1} < X_j \leq a_i, 1 \leq j \leq n\}$, $i = 1, \dots, g$. The likelihood function is then defined as

$$L_1 = \frac{n!}{\prod_{i=1}^g n_i!} \prod_{i=1}^g \left(\int_{a_i}^{a_{i-1}} f_X(x) d\mu(x) \right)^{n_i},$$

where f_X is the density of X with respect to Lebesgue measure μ . Hence

$$L_1 \propto \prod_{i=1}^g (\bar{F}_X(a_i) - \bar{F}_X(a_{i-1}))^{n_i}.$$

Then setting $\bar{F}_X(x)$ equal to $Cx^{-\alpha}$ for $x \geq a_k \geq D$, we consider the conditional likelihood function $L(\alpha|n_1, \dots, n_k)$ proportional to

$$L_k(\alpha) = \prod_{i=1}^k \left(\frac{\bar{F}_X(a_i) - \bar{F}_X(a_{i-1})}{\bar{F}_X(a_k)} \right)^{n_i} = \prod_{i=1}^k \left(\frac{a_i^{-\alpha} - a_{i-1}^{-\alpha}}{a_k^{-\alpha}} \right)^{n_i}. \quad (3.8)$$

The proposed tail-index estimator is given by

$$G_k := \arg \max L_k(\alpha) \quad (3.9)$$

where $k \in \{2, 3, \dots, g\}$. The lemma below shows that G_k exists and is a unique maximum likelihood estimator for α . As a result, one is able to obtain maximum likelihood estimates for tail probabilities and mean excess loss by using the invariance property of maximum likelihood estimators. These formulas are given in Section 6.

Lemma 3.3.0.1 *G_k in (3.9) exists and is unique.*

Proof

Define $b_i := \log(a_i/a_k)$ for $i = 1, \dots, k$ and $u_i := a_i/a_{i-1}$ for $i = 2, \dots, k$. Using equation (3.8), consider the log-likelihood function

$$\log L_k(\alpha) = \alpha n_1 \log(a_k/a_1) - \sum_{i=2}^k n_i \log \left(\frac{a_i^{-\alpha} - a_{i-1}^{-\alpha}}{a_k^{-\alpha}} \right).$$

Then it is easy to show using calculus that

$$\frac{\partial \log L_k(\alpha)}{\partial \alpha} = -n_1 b_1 - \sum_{i=2}^k n_i \left(\frac{b_i}{1 - u_i^\alpha} + \frac{b_{i-1}}{1 - u_i^{-\alpha}} \right).$$

Noting that $u_i < 1$ for each i and $b_i > 0$ for $i \geq 2$, we have

$$\frac{\partial \log L_k(\alpha)}{\partial \alpha} \longrightarrow \begin{cases} -\sum_{i=1}^k n_i b_i < 0, & \alpha \uparrow +\infty, \\ +\infty, & \alpha \downarrow 0. \end{cases}$$

The result follows by noting that $b_i > b_{i-1}$ implies

$$\frac{\partial^2 \log L_k(\alpha)}{\partial \alpha^2} = \frac{(b_i - b_{i-1}) \log u_i}{(2 \sinh(\alpha \log(u_i)/2))^2} < 0.$$

□

3.4. Performance Assessment

In this section, we conduct a simulation to study the performance of the proposed tail estimator G_k . The two key questions guiding the design of the simulation are: (1) what is the efficiency lost due to the use of partitioned data? and (2) how robust is the proposed estimator with respect to model misspecification?

Specifically, m samples of size n are generated from a distribution $F(x)$ with the mean $\mu < \infty$, standard deviation σ and $x \geq 0$. The domain of $F(x)$, \mathbb{R}^+ , is partitioned into g non-overlapping intervals, I_1, \dots, I_g . That is, $I_i \cap I_j$ for $1 \leq i \neq j \leq g$ and $\mathbb{R}^+ = \cup_{i=1}^g I_i$. The individual observations in each sample are then grouped with respect to the partition, and frequencies n_i in each interval, $i = 1, \dots, g$, are recorded. In this article, we report the simulation results obtained from using $m = 1000$ (sample), $n = 1000$ (observations), $g = 15$ (intervals), and the partition $I_i = (F^{-1}(p_i), F^{-1}(p_{i-1}))$, where $\{p_j\}_0^{15} = \{1.00, 0.995, 0.99, 0.98, 0.975, 0.95, 0.90(0.1)0.00\}$ for $i = 1, 2, \dots, g$, and $F^{-1}(p) = \inf\{x : F(x) \geq p\}$. We consider four distributions commonly used in modeling insurance losses. They include the Pareto with a parameter α , generalized Pareto with parameters γ and σ , Burr with parameters λ , θ , and τ , and the half T distributions with

TABLE 3.1: Tail index parameters and mean excess functions for selected distributions.

| Distribution | $\bar{F}_X(x) = 1 - F_X(x)$ | Parameters | $e(u)^a$ | Tail Index |
|--------------|---|-----------------------------|--|---------------|
| Pareto | $\left(\frac{D}{x}\right)^\alpha 1_{(D,\infty)}(x)$ | $D, \alpha > 0$ | $\frac{u}{\alpha-1}$, for $\alpha > 1$ | α |
| GPD | $\left(1 + \frac{\gamma}{\sigma}x\right)^{-\frac{1}{\gamma}} 1_{(0,\infty)}(x)$ | $\gamma, \sigma > 0$ | $\frac{\sigma+\gamma u}{1-\gamma}$, for $\gamma^{-1} > 1$ | γ^{-1} |
| Burr | $\left(\frac{\lambda}{\lambda+x^\tau}\right)^\theta 1_{(0,\infty)}(x)$ | $\lambda, \tau, \theta > 0$ | $\frac{u}{\alpha\tau-1}(1+o(1))$, for $\alpha\tau > 1$ | $\alpha\tau$ |
| Half-T | $\frac{2\Gamma(\frac{\phi+1}{2})}{\sqrt{\phi\pi}\Gamma(\phi/2)} \int_x^\infty \left(1 + \frac{y^2}{\phi}\right)^{-\frac{\phi+1}{2}} dy 1_{(0,\infty)}(x)$ | $\phi > 0$ | $\frac{u}{\phi-1}(1+o(1))$, for $\phi > 1$ | ϕ |

^aThe asymptotic relations are to be understood for $u \rightarrow \infty$.

degrees of freedom ϕ . The parameterizations of these distributions are given in Table 3.1.

With simulated data in two different formats, the exact values as well as in partitioned form, we compare the performance of the proposed estimator G_k using frequencies in the intervals I_i where $\inf I_i \geq D$ to that of the Hill estimator using all $x_i \geq D$, as well as to that of the maximum likelihood estimator using all frequencies n_i or all x_i . In Figures 3.1(a)–3.1(d), we report the loss in efficiency due to the use of partitioned data. The Hill estimates for α in the j^{th} box-plot, from left to right, are calculated using the largest $n_1 + \dots + n_j$ order statistics. The estimates from the proposed estimator in the j^{th} box-plot, from left to right, are calculated using (3.9) with $k = j + 1$, for $j = 1, 2, \dots, 14$. We notice that in Figures 3.1(a)–3.1(d) the proposed estimator behaves similar to the Hill estimator. In addition, we take the tail estimates that comprise each box-plot to calculate the root mean squared error (RMSE). That is, for the j^{th} box-plot, $\text{RMSE}_j = m^{-1} \sum_{i=1}^m (\hat{\alpha}_{ji} - \alpha)^2$, where $m = 1000$, the true tail index $\alpha = 1.50$, and $\hat{\alpha}_{ji}$ represents the i^{th} tail index estimate in the j^{th} box-plot. The dash line in each panel represents the true tail index parameter value. To quantify the loss of efficiency, we further define efficiency (EFF) as the ratio of RMSE_j obtained from the proposed estimator to RMSE_j obtained from the Hill estimator. The results are reported in Table 3.2.

To examine the robustness of the proposed estimator against model misspecification, we compare the proposed estimator using frequencies in top 5 and 6 intervals, which correspond to the 90th and 80th percentiles of the true underlying distribution, to four maximum likelihood (ML) estimators using all 15 frequencies n_1, \dots, n_{15} . These four ML estimators

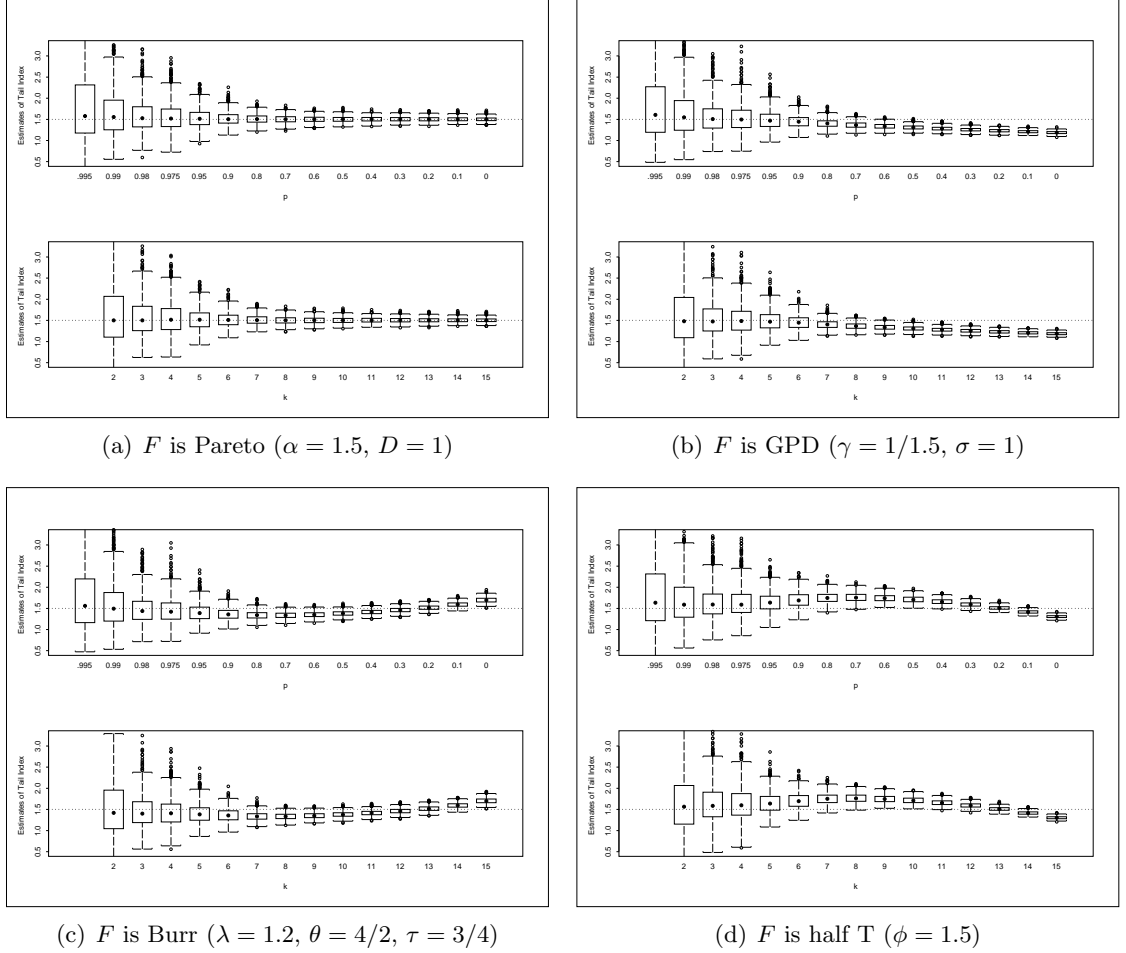
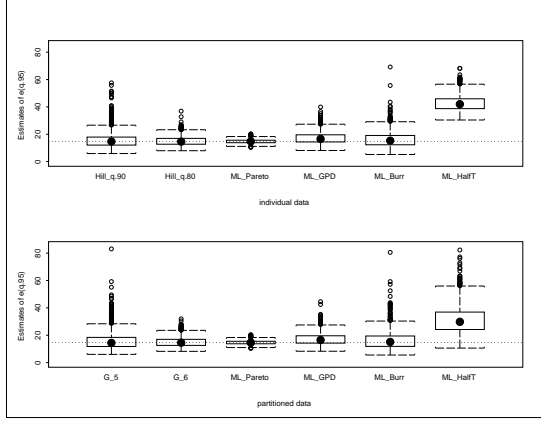


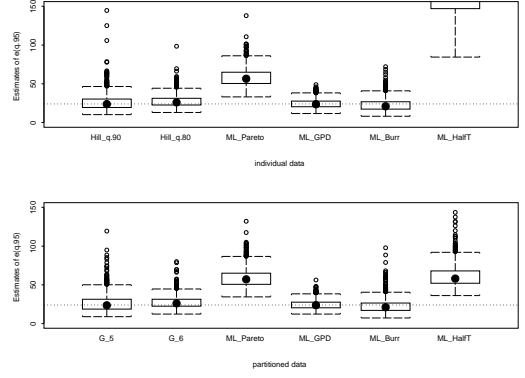
FIGURE 3.1: Performance of Hill (top) and G_k (bottom) estimators for underlying distributions (see Table 3.1) each with a true tail index of 1.5. Hill estimates use all order statistics above $F^{-1}(p)$ where F is the distribution function of the underlying distribution. Tail index estimates using grouped data are found using equation (3.9) for the given number of upper interval counts k . Sample size = number of replications = 1000.

differ in the assumed underlying distributions. They include Pareto (ML_Pareto), generalized Pareto (ML_GPD), Burr (ML_Burr), and half T (ML_T). Following our simulation design, it allows one of the four ML estimates to be the target estimate since this particular estimate is obtained by assuming the correct underlying distribution and by using the entire sample (all 15 frequencies) in estimation. The performance of the Hill estimator using observations above the 90th and 80th percentiles of the true distribution is also

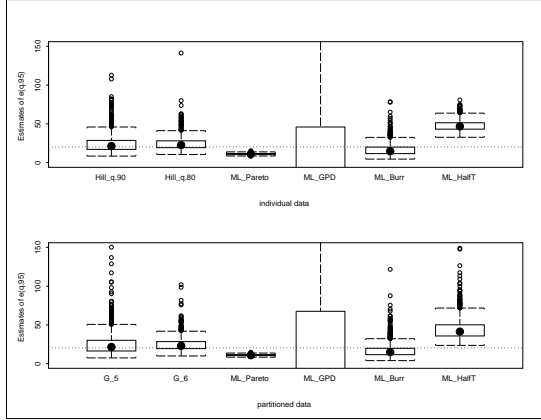
compared to those of the four similarly defined ML estimators that use the entire sample in estimation. With the tail estimates, we then calculate the expected loss exceeding the 95th percentile of the true distribution, $e(q_{.95}) = E\{X - q_{.95} | X > q_{.95}\}$. The resulting



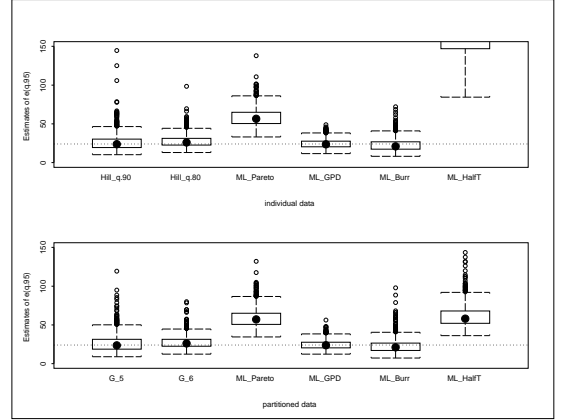
(a) F is Pareto ($\alpha = 1.5$, $D = 1$)



(b) F is GPD ($\gamma = 1/1.5$, $\sigma = 1$)



(c) F is Burr ($\lambda = 1.2$, $\theta = 4/2$, $\tau = 3/4$)



(d) F is half T ($\phi = 1.5$)

FIGURE 3.2: Estimation of Mean Excess Value $e(q_{.95})$. ML estimates are calculated under the assumption of the specified distributions (see Table 3.1). Each distribution F has a tail index of 1.5. The top plot uses all data, and the bottom plot uses grouped data. The Hill $_{q.90}$ and Hill $_{q.80}$ use all order statistics larger than $q_{.90} = F^{-1}(.90)$ and $q_{.80} = F^{-1}(.80)$. The G_5 and G_6 use the counts from top 5 and 6 intervals. Sample size = number of replications = 1000.

expected losses are reported in Figures 3.2(a) – 3.2(d). In addition, we quantify these figures by calculating RMSE and EFF (see Table 3.3). Note that EFF in this table is

defined as the ratio of RMSE of an estimator to that of the ML estimator that assumes the correct underlying distribution. Hence, if the true underlying distribution is Pareto, then $EFF = 1$ for ML_Pareto.

3.5. Discussion on Simulation Results

The simulation conducted in the previous section illustrates the loss of efficiency in using partitioned data. There is no doubt that efficiency is lost with the use of partitioned data simply because fewer data points are used in maximizing the likelihood function. This is evident from those box-plots in the far left in Figures 3.1(a) – 3.1(d) and from the EFF measures in the first few columns in Table 3.2 when only observations exceeding the 95th percentile are used in estimation. For example, as shown in Table 3.2, when the underlying distribution is Pareto, the RMSE for the Hill estimator using observations exceeding the 99th percentile and the RMSE for the proposed estimator using the frequencies from the top two intervals are 0.90 and 3.49, respectively, giving $EFF = 3.89$. This implies that parameter estimation error, measured in RMSE, can be 3.89 times higher with the use of partitioned data than with the use of individual data. However, the amount of error quickly diminishes. With only the top three frequencies (N_1 , N_2 , and N_3) in use, the EFF is below 1.20 for all four distributions. Using the top five frequencies or more, the EFF never exceeds 1.10 and quickly approaches 1.01. The parameter estimation error between the use of partitioned data and of individual data becomes negligible.

TABLE 3.2: Loss of Efficiency with the Use of Partitioned Data.

| Threshold D used in the Hill estimator | $q_{.99}$ | $q_{.98}$ | $q_{.975}$ | $q_{.95}$ | $q_{.90}$ | $q_{.80}$ | $q_{.70}$ | $q_{.60}$ | $q_{.50}$ | $q_{.40}$ | $q_{.30}$ | $q_{.20}$ | $q_{.10}$ | $q_{q,00}$ |
|--|-----------|-----------|------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|------------|
| No. of top intervals k used in G_k | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| True distribution: Pareto | | | | | | | | | | | | | | |
| Cutoff D | 21.54 | 13.57 | 11.70 | 7.37 | 4.64 | 2.92 | 2.23 | 1.84 | 1.59 | 1.41 | 1.27 | 1.16 | 1.07 | 1.00 |
| Hill | 0.90 | 0.43 | 0.35 | 0.22 | 0.16 | 0.11 | 0.09 | 0.08 | 0.07 | 0.06 | 0.06 | 0.05 | 0.05 | 0.05 |
| G_k | 3.49 | 0.49 | 0.39 | 0.24 | 0.17 | 0.11 | 0.09 | 0.08 | 0.07 | 0.06 | 0.06 | 0.06 | 0.05 | 0.05 |
| Efficiency | 3.89 | 1.14 | 1.11 | 1.08 | 1.05 | 1.04 | 1.03 | 1.02 | 1.02 | 1.02 | 1.02 | 1.01 | 1.01 | 1.01 |
| True distribution: generalized Pareto | | | | | | | | | | | | | | |
| Cutoff D | 31.82 | 19.86 | 17.04 | 10.55 | 6.46 | 3.89 | 2.85 | 2.26 | 1.88 | 1.61 | 1.40 | 1.24 | 1.11 | 1.00 |
| Hill | 0.91 | 0.38 | 0.32 | 0.22 | 0.16 | 0.14 | 0.15 | 0.17 | 0.20 | 0.22 | 0.25 | 0.27 | 0.29 | 0.32 |
| G_k | 3.46 | 0.44 | 0.36 | 0.24 | 0.17 | 0.14 | 0.15 | 0.17 | 0.20 | 0.22 | 0.25 | 0.27 | 0.29 | 0.32 |
| Efficiency | 3.82 | 1.16 | 1.10 | 1.09 | 1.06 | 1.01 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| True distribution: Burr | | | | | | | | | | | | | | |
| Cutoff D | 24.87 | 15.12 | 12.86 | 7.70 | 4.57 | 2.69 | 1.99 | 1.62 | 1.39 | 1.25 | 1.14 | 1.07 | 1.03 | 1.00 |
| Hill | 0.85 | 0.36 | 0.31 | 0.23 | 0.19 | 0.19 | 0.18 | 0.16 | 0.13 | 0.10 | 0.07 | 0.06 | 0.11 | 0.21 |
| G_k | 3.42 | 0.42 | 0.34 | 0.25 | 0.20 | 0.19 | 0.18 | 0.16 | 0.13 | 0.10 | 0.07 | 0.06 | 0.11 | 0.21 |
| Efficiency | 4.01 | 1.17 | 1.11 | 1.09 | 1.04 | 1.01 | 1.01 | 1.01 | 1.00 | 1.00 | 1.00 | 1.02 | 1.03 | 1.02 |
| True distribution: Half T | | | | | | | | | | | | | | |
| Cutoff D | 18.82 | 12.20 | 10.64 | 7.02 | 4.71 | 3.20 | 2.55 | 2.15 | 1.87 | 1.65 | 1.47 | 1.30 | 1.15 | 1.00 |
| Hill | 1.57 | 0.45 | 0.40 | 0.28 | 0.27 | 0.28 | 0.28 | 0.26 | 0.22 | 0.17 | 0.11 | 0.05 | 0.09 | 0.20 |
| G_k | 3.86 | 0.52 | 0.43 | 0.30 | 0.28 | 0.28 | 0.28 | 0.26 | 0.23 | 0.17 | 0.11 | 0.05 | 0.09 | 0.20 |
| Efficiency | 2.47 | 1.13 | 1.09 | 1.05 | 1.03 | 1.02 | 1.02 | 1.01 | 1.01 | 1.01 | 1.00 | 1.00 | 1.01 | 1.01 |

TABLE 3.3: Robustness of the Proposed Estimator Against the Underlying Distribution.

| | True $F(x)$ | RMSE | | | | | |
|------------------|-------------|-----------------|-----------------|-----------|----------|---------|----------|
| | | Hill- $q_{.90}$ | Hill- $q_{.80}$ | ML_Pareto | ML_GPD | ML_Burr | ML_halfT |
| Individual data | Pareto | 6.37 | 3.49 | 1.45 | 4.81 | 6.10 | 18.59 |
| | GPD | 11.39 | 8.61 | 36.47 | 5.79 | 8.39 | 491.45 |
| | Burr | 25.47 | 10.00 | 9.26 | 4425.52 | 8.84 | 18.65 |
| | Half T | 17.96 | 18.84 | 5.76 | 21.97 | 20.02 | 4.82 |
| Partitioned data | | G_5 | G_6 | ML_Pareto | ML_GPD | ML_Burr | ML_halfT |
| | Pareto | 7.02 | 3.63 | 1.47 | 5.04 | 7.25 | 20.23 |
| | GPD | 17.49 | 8.79 | 37.04 | 5.98 | 9.64 | 25.47 |
| | Burr | 40.12 | 10.33 | 9.33 | 42910.74 | 11.66 | 20.72 |
| | Half T | 18.03 | 18.90 | 5.67 | 22.20 | 19.26 | 12.60 |
| | True $F(x)$ | Efficiency | | | | | |
| | | Hill- $q_{.90}$ | Hill- $q_{.80}$ | ML_Pareto | ML_GPD | ML_Burr | ML_halfT |
| Individual data | Pareto | 4.40 | 2.41 | 1.00 | 3.32 | 4.22 | 12.85 |
| | GPD | 1.97 | 1.49 | 6.30 | 1.00 | 1.45 | 84.86 |
| | Burr | 2.88 | 1.13 | 1.05 | 500.56 | 1.00 | 2.11 |
| | Half T | 3.73 | 3.91 | 1.19 | 4.56 | 4.15 | 1.00 |
| Partitioned data | | G_5 | G_6 | ML_Pareto | ML_GPD | ML_Burr | ML_halfT |
| | Pareto | 4.79 | 2.48 | 1.00 | 3.44 | 4.94 | 13.80 |
| | GPD | 2.93 | 1.47 | 6.20 | 1.00 | 1.61 | 4.26 |
| | Burr | 3.44 | 0.89 | 0.80 | 3680.90 | 1.00 | 1.78 |
| | Half T | 1.43 | 1.50 | 0.45 | 1.76 | 1.53 | 1.00 |

Figures 3.1(a) – 3.1(d) also reveal a typical problem in tail index estimation. Taking only few data points in estimation, the resulting estimates exhibit large variance; whereas taking more data points than necessary, the bias of the estimates seems evident. This variance-bias tradeoff suggests the development of a threshold selection process to determine a threshold above which the assumed Pareto functional form holds. In other words, we should not include any data points that are below the threshold in estimation to avoid bias because the assumed Pareto form is no longer valid. In addition to the diagnostic plot approach described in the next section, we may also consider an analytic approach to selecting the threshold for a given sample. We may start with the frequencies N_1 and N_2 in the first two intervals I_1 and I_2 , and sequentially include frequencies in the adjacent intervals by testing whether the assumed Pareto form holds. We could perhaps make use of the fact that, conditional on $\sum_{i=1}^k N_i = \sum_{i=1}^k n_i$, $N_j \sim \text{Binomial}(\sum_{i=1}^k n_i, p_{jk}(\alpha))$, where $p_{jk}(\alpha) = (a_j^{-\alpha} - a_{j-1}^{-\alpha}) / a_k^{-\alpha}$.

If the underlying distribution is known, then the ML estimator is a common choice for parameter estimation. The ML estimate and the quantities derived from the estimate, e.g., the mean excess value $e(u)$, possess desirable statistical properties. However, the true underlying distribution is typically unknown in practice, and the penalty of model misspecification and possibly subsequent misinformed decisions may not be negligible. Our simulation results shown in Figures 3.2(a) – 3.2(d), and in Table 3.3 illustrate the robustness of our proposed estimator and the penalty of model misspecification. It is clear from Table 3.1 that a reliable estimate of the tail index is crucial for estimating the mean

excess function $e(u)$. The estimation error of $e(u)$ can be substantial without a reliable tail index estimator. For example, as reported in Table 3.3, when the true distribution is Pareto, the estimation error of $e(u)$, measured as RMSE, for the four ML estimators using individual data and partitioned data ranges from 1.45 to 18.59, and from 1.47 to 20.23, respectively. ML_Pareto, not surprisingly, has the lowest RMSE because it assumes the correct underlying distribution and utilizes the entire sample. However, if the distribution is mistakenly assumed, then the RMSE can be 3, 4 or even 13 times higher than that of ML_Pareto. In contrast, the RMSEs of the proposed estimator and the Hill estimator, despite using only a fraction of the data, stay relatively close to the best RMSE across all four assumed distributions, providing the robustness against model misspecification.

Table 3.3 also highlights a problem often encountered in practice: the ML algorithm may not converge properly leading to abnormal estimates. This is evident from the ML_GPD column where the ML algorithm did not converge in several iterations resulting in insensible estimates, and thus, large RMSE.

Finally, the Hill and G_k estimators largely underestimate $e(q_{.95})$ when the true underlying distribution is half T (Figure 3.2(d).) This is the result of the variance-bias tradeoff previously discussed. By using frequencies in the top 5 or 6 intervals, we have taken data from the area of distribution that the Pareto tail approximation does not hold. Once again, a threshold selection method is necessary to identify the optimal number k of frequencies to be used in G_k .

3.6. Applications to Insurance

In this section we apply the proposed tail index estimator to actual insurance data available only in a partitioned form. The observed losses, summarized in Table 3.4, are taken from Hogg and Klugman (1984) and consist of Homeowners 02 policies in California during accident year 1977 supplied by the Insurance Services Office (ISO). Losses were developed to 27 months and include only policies with a \$100 deductible.

TABLE 3.4: Homeowners Physical Damage

| j | a_j | a_{j-1} | Fire | | |
|------------|-------|-----------|------------------|---------------|--------------------|
| | | | $1 - F_n(a_j)^a$ | \bar{x}_j^b | $\hat{\alpha}_j^c$ |
| 1 | 50100 | ∞ | 1.21 | 78278 | NA |
| 2 | 25100 | 50100 | 3.03 | 35486 | 1.3286 |
| 3 | 10100 | 25100 | 5.83 | 16419 | 0.8779 |
| 4 | 5100 | 10100 | 9.00 | 7135 | 0.759 |
| 5 | 1100 | 5100 | 30.85 | 2256 | 0.7902 |
| 6 | 850 | 1100 | 37.99 | 974 | 0.7938 |
| 7 | 600 | 850 | 49.65 | 715 | 0.7873 |
| 8 | 500 | 600 | 57.55 | 555 | 0.7905 |
| 9 | 400 | 500 | 66.68 | 452 | 0.7684 |
| 10 | 350 | 400 | 71.91 | 378 | 0.7478 |
| 11 | 300 | 350 | 77.70 | 328 | 0.7203 |
| 12 | 250 | 300 | 83.69 | 278 | 0.6812 |
| 13 | 211 | 250 | 88.64 | 233 | 0.6435 |
| 14 | 200 | 211 | 89.90 | 207 | 0.6303 |
| 15 | 175 | 200 | 93.46 | 191 | 0.6026 |
| 16 | 156 | 175 | 95.61 | 167 | 0.5753 |
| 17 | 150 | 156 | 96.11 | 154 | 0.5653 |
| 18 | 125 | 150 | 98.92 | 141 | 0.5258 |
| 19 | 100 | 125 | 100.00 | 117 | 0.4743 |
| $n = 7534$ | | | | | |

^aProportion of losses observed greater than a_j (given as a percentage).

^bAverage of losses between a_j and a_{j-1} .

^cEstimator given in equation (3.9) using $k = j$.

To determine the threshold above which to fit the Pareto tail and estimate the tail-index, we look for a range in which the α estimates are stable. We use a plot similar to the Hill plot (see, for example, Embrichts et al (1997) and Drees et al (2000)), but modify it to be applicable for partitioned losses. Under our general framework, we consider the plot

$$\{(k, G_k) : k = 2, \dots, g\}, \quad (3.10)$$

where k is the number of top groups used to find G_k , and look for a range of k values where the plot is approximately linear. This plot is given in Figure 3.3 for the above insurance example. Notice that the plot is roughly linear for thresholds between 500 and 1100 (see also Table 3.4). We use $a_k := 500$ ($k = 8$) as the threshold and obtain $G_k = 0.7905$. This tail index suggests no finite mean for the loss distribution.

Next, we consider some important quantities in modeling large insurance claims such as, extreme tail probabilities, extreme quantiles, mean excess loss, given that losses are available only in partitioned form. Under the setup described in Section 3, $\bar{F}(x) = P(X > x)$ can be approximated by

$$\widehat{\bar{F}}(x) = \begin{cases} \bar{F}_n(a_k) (x/a_k)^{-G_k} & \text{if } x > a_k \\ \bar{F}_n(x) & \text{if } x \leq a_k, \end{cases} \quad (3.11)$$

where F_n is the empirical d.f. for the losses X_1, \dots, X_n . In Figure 3.4 this approximation is illustrated for the above Fire loss data with $x > a_k = 500$. Notice how closely the fitted tail probabilities are to the emperical tail probabilities.

Similarly, one can also approximate the conditional tail probability $P(X > x | X > a_k)$ by $(x/a_k)^{-G_k}$. An extreme quantile of the loss distribution, q_p , is defined by the relationship

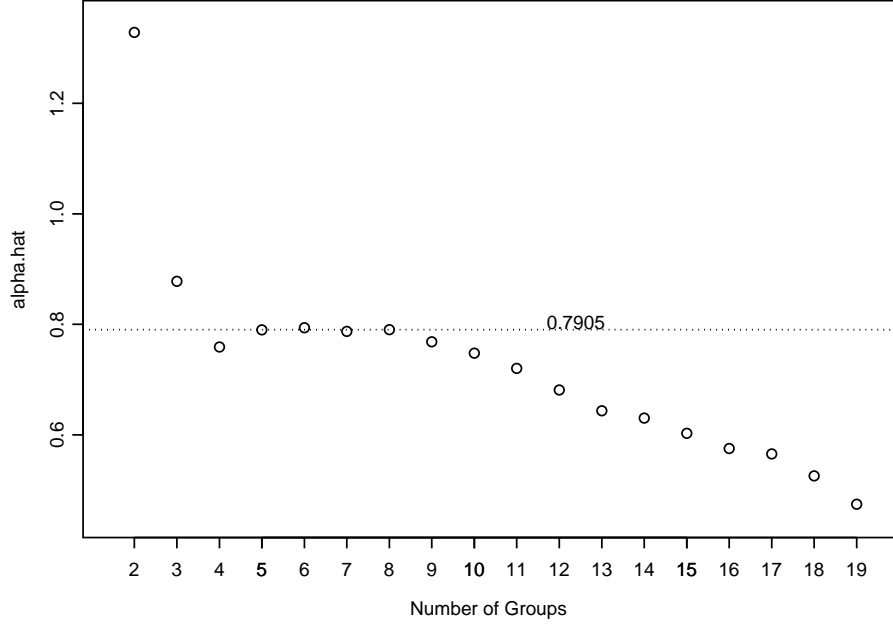


FIGURE 3.3: Tail Index Estimation for Fire Loss Data. The estimates for α using equation (3.9) are stable in the range $5 \leq k \leq 8$. This suggests to choose the cutoff $a_8 = 500$ as the threshold and to use the observed counts in top 8 intervals in equation (3.9).

$\bar{F}(q_p) = 1 - p$ where p is close to 1 (say, $F_n(a_k) < p < 1$). Setting $\widehat{\bar{F}}(x)$ equal to $1 - p$ and solving for q_p in (3.11) yields the following estimate for the extreme quantile q_p

$$\hat{q}_p = a_k \left(\frac{1 - p}{\bar{F}_n(a_k)} \right)^{-1/G_k}. \quad (3.12)$$

As an example, we estimate the .99 quantile to be $\hat{q}_{.99} = \$57,315$ using the above Fire loss data. The mean excess loss above a high threshold is important in premium determination and is given by $e(u) = E\{X - u | X > u\}$. For $u > a_k$, the mean excess loss can be approximated by

$$\hat{e}(u) = \frac{u}{G_k - 1}, \quad (3.13)$$

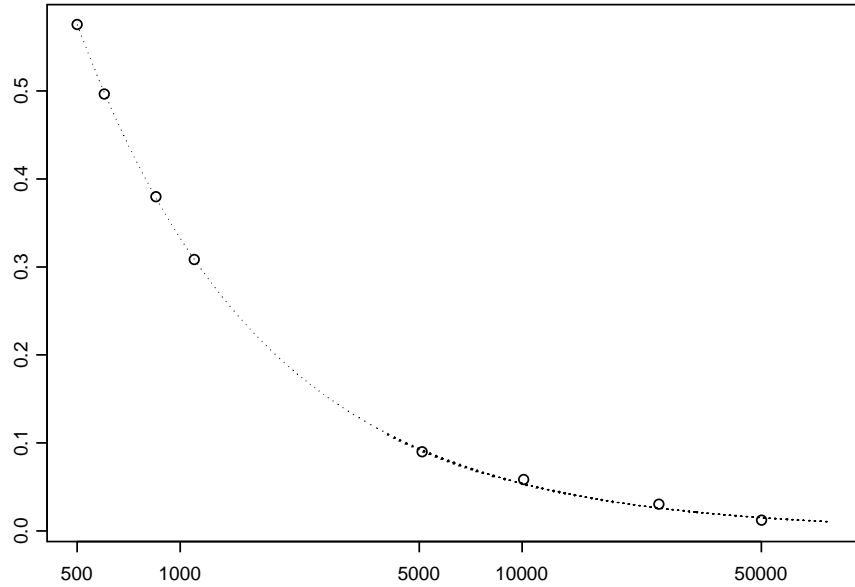


FIGURE 3.4: Comparison of Empirical and Fitted Tail Probabilities for Fire Loss Data. $\bar{F}_n(x)$ is given by open circles and $\widehat{\bar{F}}(x)$ by the dashed line where $\hat{\alpha} = 0.7905$ and $a_k = 500$. Note that the x axis is on log scale.

for $G_k > 1$. In this example, however, $\hat{e}(u)$ is not available because that $G_k \leq 1$.

3.7. Summary and Conclusion

It has been shown that losses for many lines of insurance posses Pareto-type tails. For this reason, tail index estimation, which is a measure of the heavy-taildness of a distribution, is an important problem for actuaries. Most estimators, however, cannot be used when loss data are available only in a partitioned form. The proposed estimator possesses the

attractive features of (1) being applicable when loss data are available only in a partitioned form, and (2) being robust with respect to a large class of distributions commonly used in modeling insurance losses. We also showed that tail index estimates can be misleading if one misspecifies the distribution when trying to fit a global density. We have demonstrated that the proposed estimator compares favorably to the Hill estimator that uses individual data, and provided an example showing its effectiveness using actual insurance loss data.

4. MODELING SEA SURFACE TEMPERATURE EXTREMES: A LOOK AT THE THERMOSTAT HYPOTHESIS

John B. Henry III

Abstract

The implications of the thermostat hypothesis for sea surface temperature (SST) are of interest to marine ecologists and have important ramifications for coral reef ecosystems. The thermostat hypothesis is considered here from an extreme value theory point of view resulting in an estimated SST upper bound approaching $31.2^{\circ} - 32.0^{\circ}\text{C}$ (depending on latitude) as time $\rightarrow \infty$. This estimate is obtained from a generalized Pareto model fit to SST data from the western Pacific warm pool and compared to estimates obtained via physical models.

4.1. Introduction

The increase in sea surface temperature (SST) extremes over the last few decades has greatly impacted many marine ecosystems (Richardson and Schoeman (2004)). In particular, increasing SST is believed to play a major role in a rise in the frequency and severity of coral reef bleaching events (Hoegh-Guldberg et al (2007), Berkelmans et al (2004)). It is thought that reefs in warmer regions, possibly with SST extremes already near an upper thermal limit, may be less exposed to SST increases than reefs in other regions (Kleypas et al (2008)). As coral reefs are among the most diverse and productive communities on earth, which have functions ranging from providing food and shelter to fish and invertebrates to protecting the shore from erosion, this question of whether or not there is some maximum SST is valuable.

4.1.1 The Thermostat Hypothesis

The famous 1991 Ramanathan-Collins “thermostat” hypothesis states that SST is largely regulated by strong negative feedback of cirrus clouds induced by deep convection, and that these mechanisms act to depress SST warming beyond a certain temperature (Ramanathan and Collins 1991). In particular, Ramanathan and Collins assert the following:

Observations made during the 1987 El Niño show that in the upper range of sea surface temperatures, the greenhouse effect increases with surface temperature at a rate which exceeds the rate at which radiation is being emitted from the surface. In response to this ‘super greenhouse effect’, highly reflective cirrus clouds are produced which act like a thermostat shielding the ocean from solar

radiation. The regulatory effect of these cirrus clouds may limit sea surface temperatures to less than 305 K (31.85°C).

This notion of an ocean thermostat has not been accepted by all. For example, Wallace (1992) argues that:

... although cirrus clouds reduce the solar insolation at the Earth's surface in regions of deep convection, they would not necessarily prevent SSTs from exceeding 305K (31.85°C) in the face of extensive greenhouse warming.

Similarly, Fu et al (1992) argue that changes in the properties of cirrus clouds do not seem to be related to changes in SSTs.

The term “thermostat hypothesis” was coined by Ramanathan and Collins to describe one of the SST feedback mechanisms: the increase in cloud cover due to SST-driven convection. However, if the ocean thermostat really exists, it is likely caused by a combination of feedback mechanisms. In particular, besides (1) the cloud-SST feedback or cloud shortwave radiative forcing, Kleypas et al (2008) discuss (2) latent heat flux or evaporation-wind-SST feedback and (3) ocean dynamics and heat transport as other main processes that have been proposed to limit open ocean SSTs. These mechanisms are included in the net heat flux at the ocean surface, given by

$$Q_N = Q_S - (Q_I + Q_H + Q_L), \quad (4.1)$$

where N , S , I , H and L refer to net heating, absorption of solar radiation by the ocean, net loss of infrared radiative energy from the ocean surface, loss of sensible heat through

turbulent fluxes at the ocean surface, and net loss of latent heat through the evaporation of water from the surface, respectively (Weare et al (1981)).

Li et al (2000) provide a conceptual two-box model that contains dynamic coupling among the Walker circulation, SST, and ocean thermocline and thermodynamic coupling, which includes shortwave and longwave cloud forcing and latent and sensible heat fluxes at the ocean surface. Their work provides a framework that combines the three SST-regulation mechanisms discussed above (cloud-SST feedback, evaporation-wind-SST feedback, and ocean dynamics and heat transport). They find that cloud shortwave radiation forcing makes the largest contribution to limiting SSTs in the WPWP, followed by the effects of surface evaporation, and ocean dynamics. They find that all three mechanisms are essential in limiting SSTs.

4.1.2 WPWP SST Data

While analysis of SST usually involves modeling means (Huang and Liu (2001), for example), few [any?] have conducted an analysis focusing their attention on SST extreme. Because information on extreme temperatures is lost when averages are taken over large time scales, daily SST records are used here to investigate the thermostat hypothesis. More specifically, daily time series records from buoys¹ located at 156°e and 8°n, 5°n, 2°n, 0°n, 2°s, and 5°s in the western Pacific warm pool (WPWP) are used. The earliest

¹ These buoys are part of the TAO-TRITON (Tropical Atmosphere Ocean - Triangle Trans Ocean Buoy Network) array. The author would like to acknowledge the TAO Project Office of NOAA/PMEL. More information can be found at <http://www.pmel.noaa.gov/tao/>.

time series records used here start in 1991 and the last measurements used here end in 2002. SST records are in units of degrees centigrade and were taken at a depth of 1m.

4.1.3 Organization of paper

Extreme value theory (EVT) methods have been commonly used in hydrology (Katz et al (2002)), meteorology (Smith (1999)), geology and seismic analysis (Caers et al (1999)), but not typically used to estimate upper bounds. A background of basic EVT, including motivation for using the generalized extreme value (GEV) distribution and/or generalized Pareto (GP) distribution is given in Section 4.2. In Section 4.3., EVT methods are described that are used here to model SST extremes. Analysis of SST extremes and results are provided in Section 4.4.. In this section an estimate for the SST upper bound is provided that is a function of space and time. Standard error calculation and model selection methods are described and diagnostic plots are provided. A discussion of results and their implications is provided in Section 4.5., along with a comparison of estimated SST upper bounds obtained physical models and the EVT model.

4.2. EVT Background

4.2.1 Method of Block Maxima

The Fisher-Tippett-Gnedenko theorem (Fisher and Tippett (1928), Gnedenko (1943)) is the basis of classical extreme value theory. The theorem provides us with the only

possible (non-degenerate) limiting distribution for normalized maxima of an independent and identically distributed (iid) sequence of random variables, the *generalized extreme value* (GEV) distribution. More specifically, for an iid sequence of random variables $\{X_k\}_1^n$, there exist sequences of constants $\{a_n\} > 0$ and $\{b_n\}$ such that

$$P\left(\frac{\max\{X_1, \dots, X_n\} - b_n}{a_n} \leq y\right) \rightarrow G(y) \text{ as } n \rightarrow \infty \quad (4.2)$$

for a non-degenerate distribution function (d.f.) G if and only if G is the GEV distribution function

$$G_\theta(y) = \begin{cases} \exp\left\{-\left[1 + \xi\left(\frac{y-\mu}{\beta}\right)^{-1/\xi}\right]\right\} & \text{if } \xi \neq 0 \\ \exp\left[-\exp\left\{-\left(\frac{y-\mu}{\beta}\right)\right\}\right] & \text{if } \xi = 0 \end{cases}, \quad \theta := (\mu, \beta, \xi) \quad (4.3)$$

defined on $\{y : 1 + \xi(y - \mu)/\beta > 0\}$ if $\xi \neq 0$ and \mathbb{R} if $\xi = 0$. Here $\mu, \xi \in \mathbb{R}$, $\beta > 0$, and \mathbb{R} denotes the real numbers. Throughout the remainder of this paper, the notation $W \sim \text{GEV}(\theta)$ will mean that the random variable W has a GEV d.f. as in (4.3).

The three parameter in (4.3) represent location (μ), scale (β), and shape (ξ). The shape parameter ξ sometimes called the *extreme value index* is the most important in characterizing the GEV distribution. In particular, the GEV distribution can take three different forms depending on the value of ξ : (1) a bounded (or Weibull) distribution if $\xi < 0$, (2) a light-tailed (or Gumbel) distribution if $\xi = 0$, and (3) a heavy-tailed (or Frechet) distribution if $\xi > 0$. Temperature readings typically correspond to the Weibull type (see Brown and Katz (1995) for example).

In many applications, the X_i represent daily phenomena (sea level, mean temperature, ect) so that $n = 365$ and $M_n := \max\{X_1, \dots, X_n\}$ represents an annual maxima. One typically

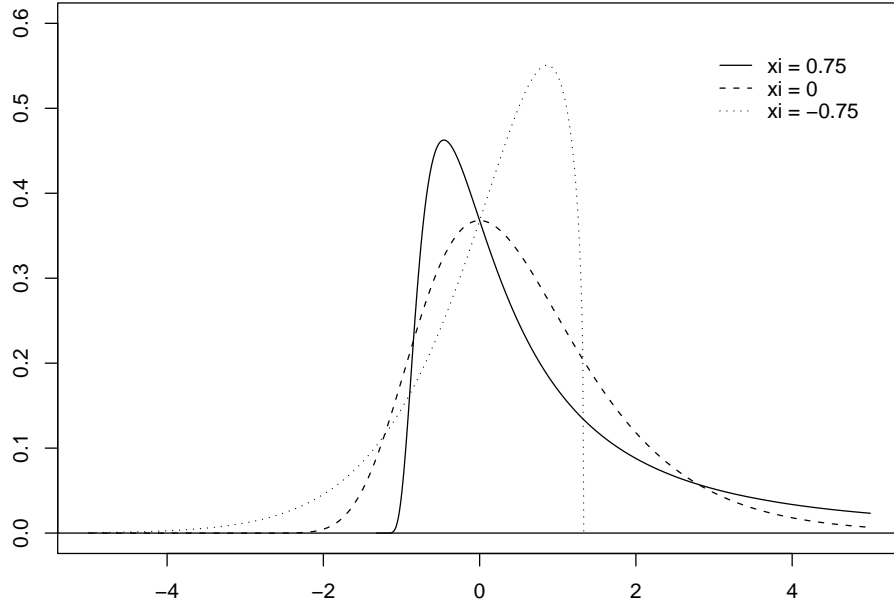


FIGURE 4.1: GEV Densities for different parameters ξ , $\mu = 0$, and $\beta = 1$ in (4.3).

works with a sequence of *block* (say, annual) maxima $M_{n,1}, M_{n,2}, \dots$ when fitting the GEV model. In practice, the problem of not knowing the normalizing constants $\{a_n\} > 0$ and $\{b_n\}$ is not troublesome due to the fact that if $(M_n - b_n)/a_n$ is approximately $\text{GEV}(\theta)$ distributed, then M_n is approximately $\text{GEV}(\theta')$ distributed, where $\theta \neq \theta'$. Since estimates of the parameters of the GEV distribution are usually required anyway, estimating θ' rather than θ provides no additional work.

4.2.1.1 Parameter Estimation

Maximum likelihood (ML) estimation is a common choice for estimating the parameters in the GEV distribution. Given a collection of iid GEV random variables $\mathbf{y} = (y_1, \dots, y_n)$, the ML estimator is given by

$$\hat{\theta} := \arg \max_{\theta \in \Theta} L(\theta; \mathbf{y}), \quad (4.4)$$

where $L(\theta; \mathbf{y}) = \prod_i g_\theta(y_i)$. Here L is the likelihood function, g is the GEV density function, and Θ is the appropriate parameter space. Because no closed form solution exists for the estimator in (4.4), numerical methods must be used. ML estimation has an advantage over many other types of estimation approaches in the way that covariates can easily be included. Including covariates can make use of relevant information and relaxes the iid assumption. See Coles (2001) and Katz et al (2005) for examples. Problems can occur in ML estimation of the GEV parameters when using small samples and/or when $\xi \leq -1/2$ (Smith (1985)). Other estimation methods can be found in Embrechts et al (1997).

4.2.2 Threshold Models

Rather than only making use of maxima, as in the previous subsection, one often is interested in observations of a phenomena exceeding some high threshold. Also, it is also more efficient to consider threshold models even if the ultimate interest is in block maxima. The generalized Pareto (GP) distribution turns out to be the analogue of the

GEV distribution in this situation. The GP d.f. is given by

$$H_{\xi,\sigma}(y) = \begin{cases} 1 - \left(1 + \frac{\xi y}{\sigma}\right)^{-1/\xi} & \text{if } \xi \neq 0 \\ 1 - \exp\{-y/\sigma\} & \text{if } \xi = 0 \end{cases} \quad (4.5)$$

where $y \in [0, \infty)$ when $\xi \geq 0$ and $y \in [0, -\sigma/\xi]$ when $\xi < 0$.

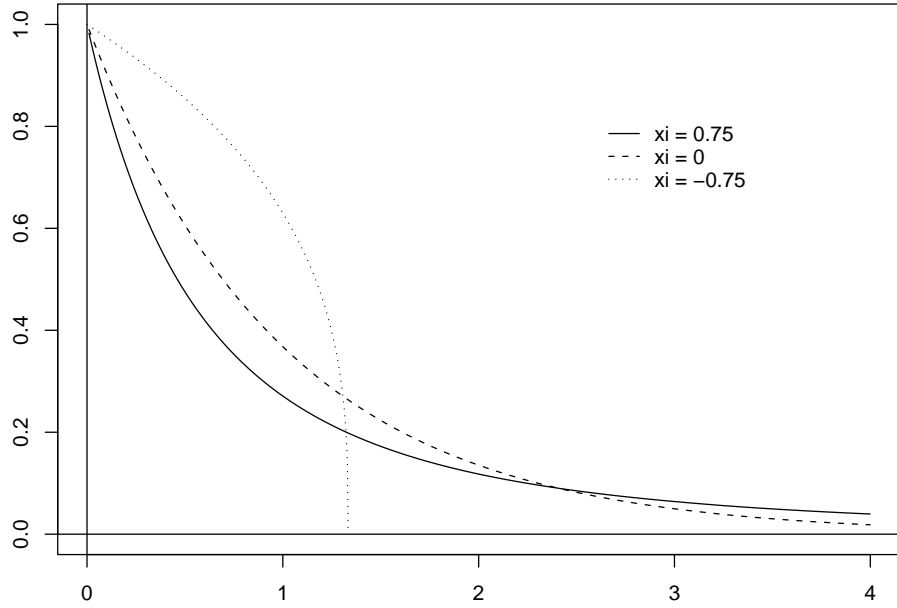


FIGURE 4.2: GP Densities for different parameters ξ and $\sigma = 1$ in (4.5).

The GP distribution is closely related to the GEV distribution. In particular, according to Pickands Theorem (Pickands (1975)), an iid sequence of random variables $\{X_k\}_1^n$, each with d.f. F , are in the maximum domain of attraction of the GEV distribution (that is, satisfy (4.2) with G as in (4.3)), with parameter $\xi \in \mathbb{R}$, if and only if

$$\lim_{u \rightarrow x_F} \sup_{0 < x < x_F - u} |P(X - u \leq x | X > u) - H_{\xi,\sigma(u)}(x)| = 0$$

for some positive function $\sigma(u)$. Here

$$x_F = \sup\{x \in \mathbb{R} : F(x) < 1\} \quad (4.6)$$

denotes the right endpoint of F . Roughly speaking, Pickands Theorem tell us that if M_n can be closely approximated by the GEV distribution with parameter ξ , then $(X - u)|X > u$ can be closely approximated by the GP distribution with the same shape parameter ξ . It is important to notice above that σ in GP distribution is dependent on the threshold u that is used.

Fitting the GP distribution is done by using the sequence $\{Y_i\}_1^{n_u}$, where $Y_i = (X_i - u)|X_i > u$ and n_u is the number of X_i greater than u . Like the GEV distribution described in Section 4.2.1, the GP distribution is flexible in the sense that it can have a power-law tail ($\xi > 0$), exponential tail ($\xi = 0$), or bounded tail ($\xi < 0$). Negative estimates of ξ that are significantly different than 0, suggest that the underlying distribution of the X_i 's has a bounded tail, in which case an estimate for the upper bound in Equation (4.6) is given by

$$\hat{x}_F = u - \hat{\sigma}_u / \hat{\xi}. \quad (4.7)$$

This is obtained by setting the d.f. in (4.5) to 1, in the case $\xi < 0$, solving for x , and replacing parameters with their estimates. This idea of an estimated upper bound will be important in examining the thermostat hypothesis using the SST data from the WPWP in Section 4.3. - 4.5.

4.2.2.1 Parameter Estimation

Because of the importance of the parameter ξ , a considerable amount of attention has been put into different methods of estimating. ML estimation can be done in the same way that was described for the GEV distribution for $\xi \in \mathbb{R}$. Hsieh (2002) provides a summary of many estimators for ξ in the case when $\xi > 0$. Other estimation methods can be considered such as method of moment estimation (Dekkers et al (1989)) if $\xi < 1/2$, and method of probability-weighted moments (Hosking et al (1985)) if $\xi < 1$. As mentioned in Section 4.2.1 however, ML estimation is usually the only approach that extends easily to models including covariates. For this reason, ML estimation is used to model SST extremes in Section 4.3.

4.2.2.2 Threshold Selection

Using the GP distribution to model excess above high thresholds requires a choice of a threshold, above which the GP assumption is reasonable. While there do exist *automatic* threshold selection algorithms, careful consideration is usually (and should be) taken when selecting a threshold. If an extremely high threshold is used, one may feel comfortable with assuming a GP distribution beyond that threshold. However, this may result in the availability of few large order statistics for parameter estimation. On the other hand, choosing a low threshold provides more order statistics to use, but can introduce bias in parameter estimates if the GP assumption is inappropriate. There are several useful diagnostic plots that can aid in choosing a threshold. For example, a plot of the sample

mean excess over thresholds u is expected to be linear in u with slope $\xi/(1 - \xi)$ when the GP assumption is valid. In particular, we expect this plot to increase if $0 < \xi < 1$, decrease if $\xi < 0$, and be roughly constant if ξ is near 0. Note that the mean excess does not exist for $\xi > 1$.

4.3. Methods

The GP distribution is used to model excess over a threshold, rather than the method of block maxima, of extreme SSTs in the WPWP. This is done, in part, to make the most use of the the limited number of years of data available. Because of the non-stationarity of the SST time series, the first steps here are to model trend and seasonality components. After removing these from the SSTs, residual exceedances over high thresholds will be assumed to follow a GP distribution. For each location, the first daily SST corresponds to $t = 1$.

4.3.1 SST Extremes: Trend and Seasonality

A trend term is fit to each location using the WPWP data. In particular, because only upper extremes are of interest here, the trend term for each location is fit only to those SST points that were above the mean SST for each location. A trend process of the form

$$M_i(t) = a_i - b_i e^{-c_i t}, \quad a_i, b_i, c_i \geq 0, \quad i = 1, \dots, 6, \quad (4.8)$$

is assumed for each of the six locations in the WPWP. Here location $i = 1, \dots, 6$ corresponds to latitude values of $\{8^\circ\text{n}, 5^\circ\text{n}, 2^\circ\text{n}, 0^\circ\text{n}, 2^\circ\text{s}, 5^\circ\text{s}\}$ and a longitude of 156°e . In the GP model defined in (4.13) latitude, s , is treated as continuous for $-5 \leq s \leq 8$.

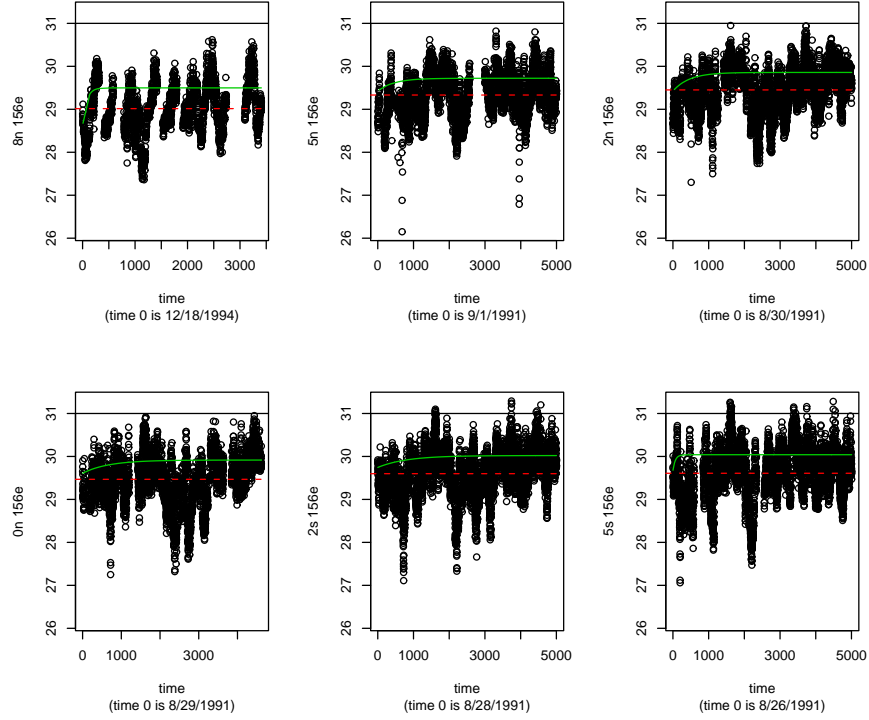


FIGURE 4.3: SST trends are estimates of equation (4.8) for each location and shown by the green lines. The means SSTs are shown by red lines, and the solid black lines are constant ($= 31^\circ$) for references.

From here on \hat{M} will denote the trend process in (4.8) with (a, b, c) replaced by $(\hat{a}, \hat{b}, \hat{c})$.

While different functional forms of M can (and were) considered, the trend given in (4.8) was chosen so that estimated upper bounds (provided $\hat{\xi} < 0$, see Section 4.4.2) approach a finite limit as $t \rightarrow \infty$.

We see in Figure 4.3 that the largest observed SSTs have occurred in the most recent

years. Notice also that the mean SST and the number of extremes increases moving from 8°n to 5°s .

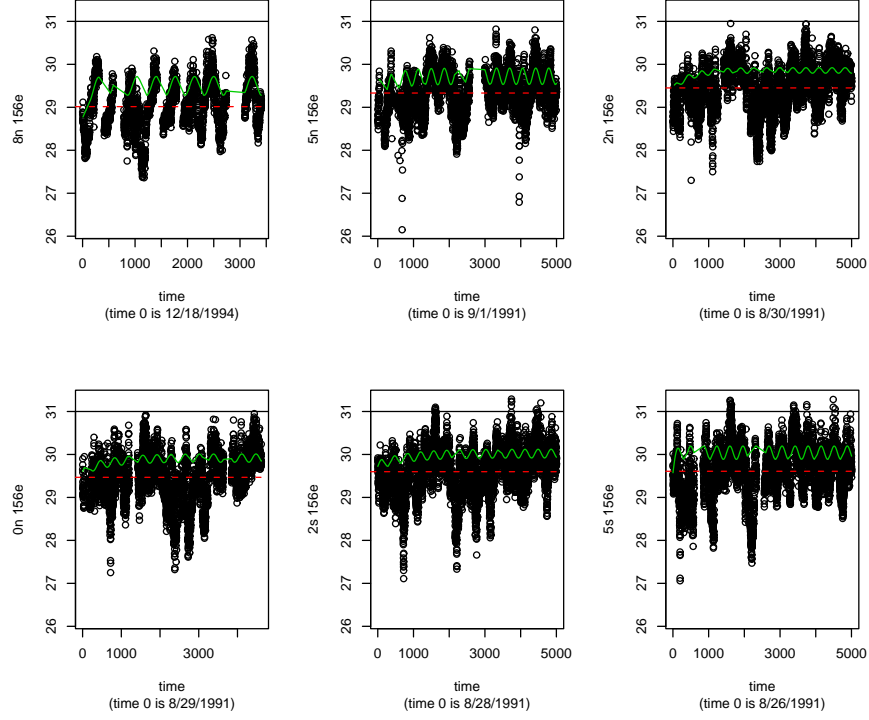


FIGURE 4.4: The SST seasonality component is given by estimates of equation (4.9). Here $\hat{M} + \hat{S}$ is shown by the green lines for each location. The means SSTs are shown by red lines, and the solid black lines are constant ($= 31^\circ$) for references.

A seasonality term of the form

$$S_i(t) = v_i \sin\left(\frac{2\pi t}{365.25}\right) + w_i \cos\left(\frac{2\pi t}{365.25}\right), \quad v_i, w_i \in \mathbb{R}, \quad i = 1, \dots, 6, \quad (4.9)$$

is also fit to each location using the WPWP data. The seasonal component is fit to each location after the estimated trend has been removed. Estimated trend and seasonality terms are shown in Table 4.1 in Section 4.4.1. In Figure 4.4 the estimated seasonality term \hat{S} (obtained by replacing (v, w) with (\hat{v}, \hat{w}) in (4.9)) is added to the estimated trend

\hat{M} for each location.

4.3.2 Residuals and Excess

Residuals are obtained by removing the trend and seasonal components at each location.

The residual SST process at each location is defined as

$$SST_i(t) - M_i(t) - S_i(t) := R_i(t). \quad (4.10)$$

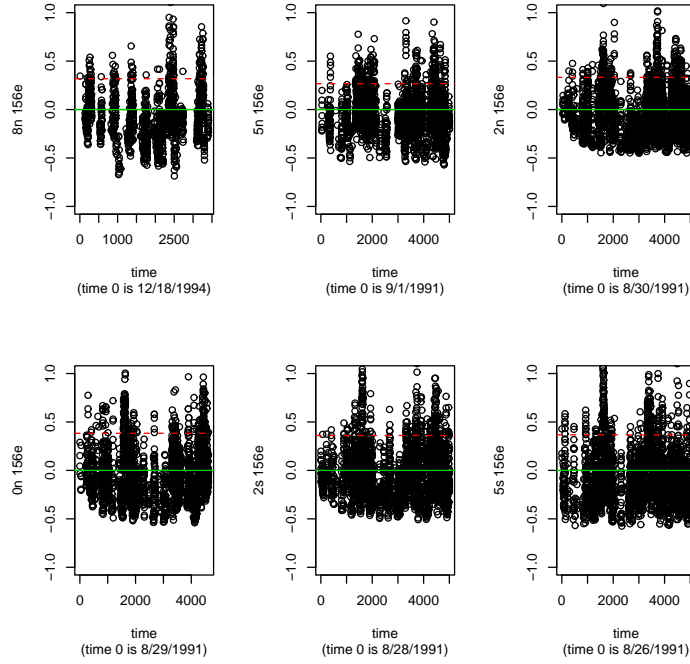


FIGURE 4.5: Residuals described in (4.10) are shown. Threshold u_i (see (4.11) and (4.12)) are shown (dashed lines) above which the GP model is fit.

The residual processes defined in (4.10) and shown in Figure 4.5 are assumed to be stationary and have mean zero. From these residuals, conditional excesses,

$$(R_i(t) - u_i) | R_i(t) > u_i := E_i(t), \quad (4.11)$$

are used in Section 4.3.3 to construct a GP model from which estimated SST upper bounds are obtained. For location i , the 0.9 sample residual quantile is used for the threshold u_i in (4.11). More specifically,

$$(u_1, \dots, u_6) = (0.318, 0.267, 0.333, 0.383, 0.361, 0.367). \quad (4.12)$$

4.3.3 GP Model for SST

With the trend and seasonal components removed from the SST processes, we assume that the excess given in Equation (4.11) satisfies

$$E(s, t) \sim \text{GP}(\xi, \sigma(s)) \quad (4.13)$$

where $\log \sigma(s) = \beta_0 + \beta_1 s$. That is, we assume that $E(s, t)$ has distribution function $H_{\sigma(s), \xi}$ where H was defined in (4.5). Here s denotes latitude values, $s \in [-5, 8]$. The location subscript i in (4.11) is dropped and the spatial differences in the excesses are modeled through the continuous scale parameter $\sigma(s)$ in the GP distribution.

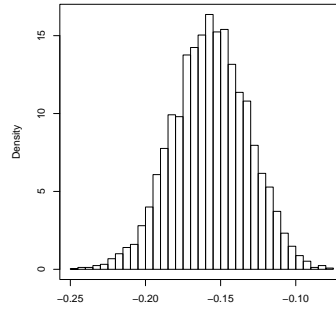
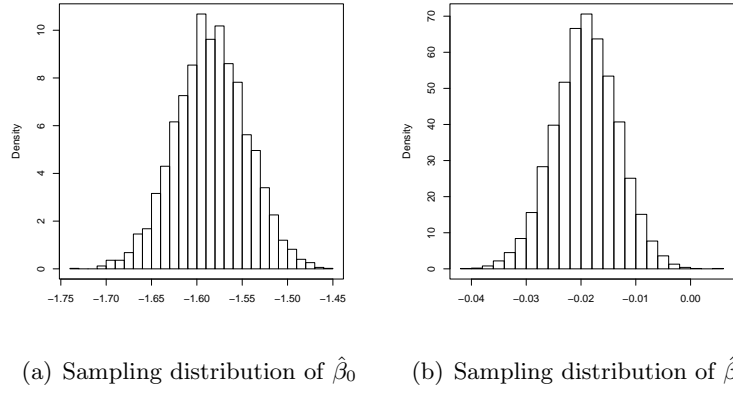
Notice here that ξ does not depend on either s or t . The assumption of a constant shape parameter ξ over space is commonly made in environmental applications. For example, Buishand (1991) assumes the distributions of rainfall at different locations have the same upper tail thickness; that is, the distributions share a common EVI ξ . Smith (1989) makes the same assumption across all basins in a regional flood frequency analysis.

As can be seen in Figure 4.3, clusters of SST extremes are not uncommon in time. This temporal dependence of extremes is still seen in the residuals (see Figure 4.5) obtained by

removing trend and seasonal components. While the common practice when performing POT modeling² is to do some kind of declustering, the suggestion of Fawcett and Walshaw (2007) is followed here, and all exceedances of a high threshold are used. The reason declustering methods are commonly used is because of the lack of independence (in time) in clusters of extremes. Fawcett and Walshaw (2007) argue that declustering can cause serious bias to occur in the estimation of GP parameters. One method of declustering is known as *runs declustering*. See Gilleland and Katz (2005) for more on runs declustering. Coles (2001) notes that results can be sensitive to the arbitrary choices made in cluster determination.

Besides the obvious temporal dependence of extreme SSTs, there is also spatial dependence that needs to be addressed. Buishand (1991) looked at modeling extreme rainfall by combining data from several sites and Smith (1989) conducted a regional flood frequency analysis. In this work, latitude is included by modeling residual excesses using the scale parameter $\sigma(s)$ in the GP distribution. Maximum likelihood estimation is used to estimate the parameters β_0 , β_1 , and ξ in (4.13), but rather than using the usual asymptotic properties of MLEs to obtain standard errors for the estimates, standard errors are obtained through resampling. See Figure 4.3.3. Katz et al (2002) Section 3.3 provides more details on standard errors, resampling, and residuals in this type of setting.

² Fitting a GP distribution to excess over high thresholds is often referred as *peaks over threshold* (POT) modeling.



(c) Sampling distribution of $\hat{\xi}$

FIGURE 4.6: Sampling distributions (from 5000 bootstrapped estimates) of $\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{\xi}$ obtained from GP model fit to excesses.

4.4. Results

4.4.1 Trend and Seasonality

Estimated trend and seasonal terms are summarized in Table 4.1. For each location, a , b , and c in (4.8) were significantly different from 0 at a 0.05 level. All but one of the coefficients in (4.9) were significantly different from 0 at a 0.05 level.

| | 8n 156e | | 5n 156e | | 2n 156e | |
|----------|----------------|--------|----------------|--------|----------------|--------|
| | estimate | SE | estimate | SE | estimate | SE |
| <i>a</i> | 29.498 | 0.0098 | 29.722 | 0.0070 | 29.858 | 0.0062 |
| <i>b</i> | 0.903 | 0.3622 | 0.296 | 0.0827 | 0.434 | 0.0649 |
| <i>c</i> | 0.013 | 0.0034 | 0.002 | 0.0007 | 0.002 | 0.0004 |
| <i>v</i> | -0.181 | 0.0116 | 0.113 | 0.0067 | 0.060 | 0.0071 |
| <i>w</i> | 0.122 | 0.0114 | 0.154 | 0.0075 | 0.023 | 0.0076 |

| | 0n 156e | | 2s 156e | | 5s 156e | |
|----------|----------------|--------|----------------|--------|----------------|--------|
| | estimate | SE | estimate | SE | estimate | SE |
| <i>a</i> | 29.916 | 0.0095 | 30.021 | 0.0086 | 30.040 | 0.0066 |
| <i>b</i> | 0.306 | 0.0572 | 0.277 | 0.0379 | 0.368 | 0.1766 |
| <i>c</i> | 0.001 | 0.0004 | 0.001 | 0.0003 | 0.015 | 0.0075 |
| <i>v</i> | -0.181 | 0.0116 | 0.113 | 0.0067 | 0.060 | 0.0071 |
| <i>w</i> | 0.122 | 0.0114 | 0.154 | 0.0075 | 0.023 | 0.0076 |

TABLE 4.1: Estimates and SEs of a, b , and c in (4.8) and v and w in (4.9).

4.4.2 Estimated Upper Bound

The GP model described in Section 4.3.3 fit to residual excess using the WPWP data results in parameter estimates in (4.13) of $(\hat{\beta}_0, \hat{\beta}_1, \hat{\xi}) = (-1.587, -0.019, -0.154)$. The sampling distributions of these estimators, as shown in Figure 4.3.3, obtained via resampling, results in 95% confidence intervals of $(-1.663, -1.509)$ for β_0 , $(-0.031, -0.008)$ for β_1 , and $(-0.207, -0.109)$ for ξ . Figure 4.7 provides a diagnostic plot for checking the fit of the model. The probability plot consists of the points

$$\left\{ \left(H_{\hat{\xi}, \hat{\sigma}(u)}(x_{(i)}), \frac{i}{n+1} \right) : i = 1, \dots, n \right\}$$

where H is the GP distribution function given in (4.5) and $x_{(1)} \leq \dots \leq x_{(n)}$ are the ordered excesses from (4.13).

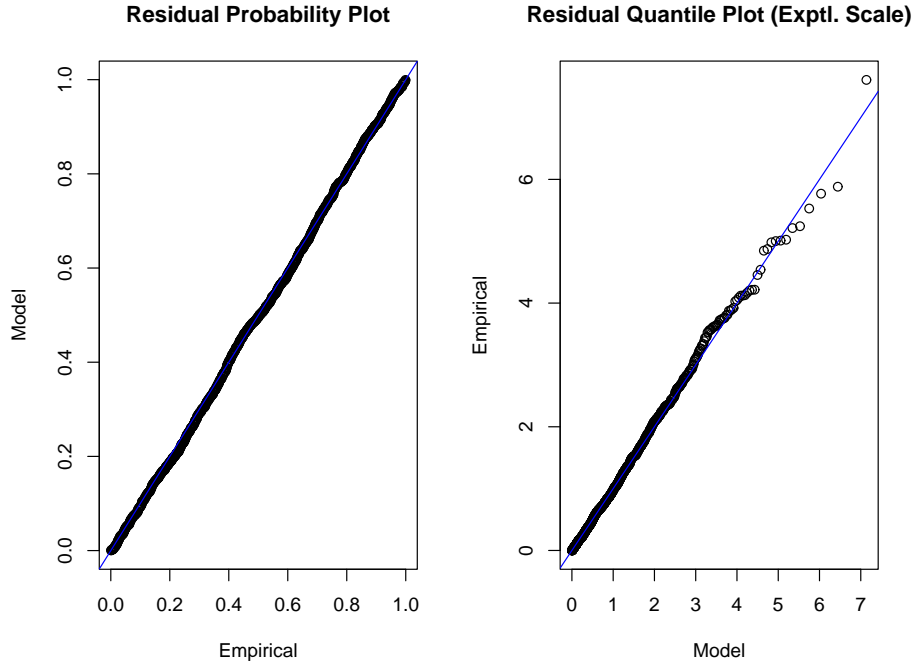


FIGURE 4.7: GP fit to WPWP data assuming (4.13) using maximum likelihood estimation.

The quantile plot consists of the points

$$\left\{ \left(H_{\hat{\xi}, \hat{\sigma}(u)}^{-1} \left(\frac{i}{n+1} \right), x_{(i)} \right) : i = 1, \dots, n \right\}$$

The probability and quantile plots show that $H_{\hat{\xi}, \hat{\sigma}(u)}$ is a reasonable model if the points lie close to the unit diagonal. This appears to be the case in Figure 4.7.

Assuming (4.13), and noting that $\hat{\xi} < 0$ is significantly different from 0, the SST upper bound has the form

$$x_F(s, t) = M + S + u - \frac{\sigma(s)}{\xi}, \quad (4.14)$$

where trend M , seasonality S , and thresholds u were given in (4.8), (4.9), and (4.12).

The estimated SST upper bound is obtained by replacing M , S , and (β_0, β_1, ξ) with their

estimates. This estimated SST upper bound is shown in Figure 4.8 varying in space (latitude) and time.

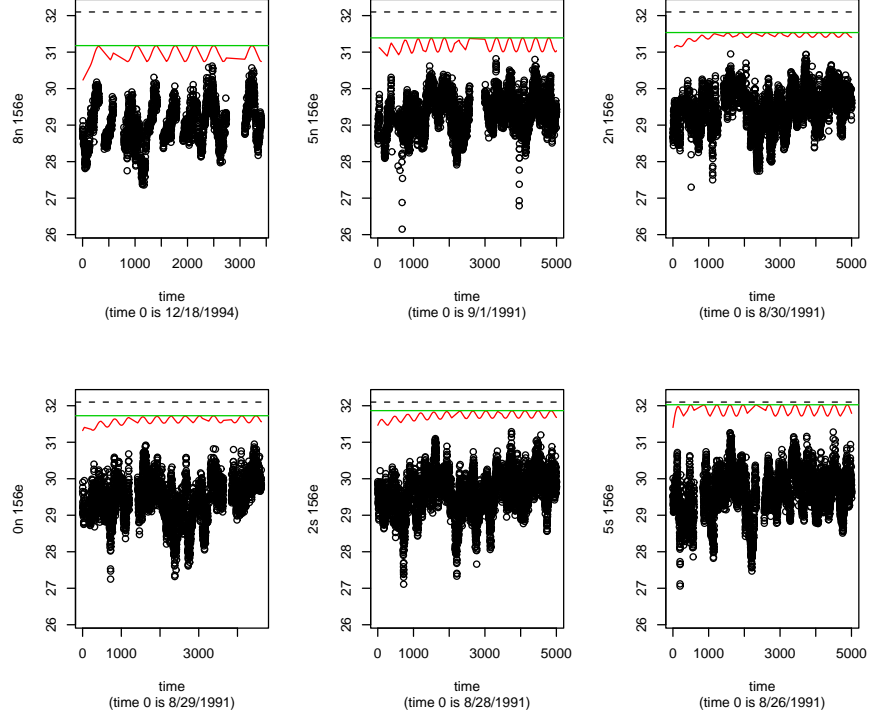


FIGURE 4.8: Estimated upper bound using (4.14) is shown in red along with the limiting upper bound from (4.15) in green. See Table 4.2. The solid black lines are constant ($= 32.1^\circ$) for references.

The limiting SST upper bound is defined here as

$$\overline{x_F(s, t)} = \limsup_t x_F(s, t), \quad (4.15)$$

becomes

$$u + a + c(v, w) - \frac{\exp\{\beta_0 + \beta_1 s\}}{\xi}$$

where $c(v, w)$ maximizes $v \sin\left(\frac{2\pi t}{365.25}\right) + w \cos\left(\frac{2\pi t}{365.25}\right)$. Note that $be^{-ct} \rightarrow 0$ as $t \rightarrow \infty$.

The estimated limiting SST upper bound, $\overline{\hat{x}_F(s, t)}$, is shown in Figure 4.8, and is constant

in time and increasing as we move south in latitude. These estimates are summarized in Table 4.2.

| Estimated Limiting SST Upper Bound | | | | | | |
|---|-------|-------|-------|-------|-------|-------|
| s | 8 | 5 | 2 | 0 | -2 | -5 |
| $\overline{\hat{x}_F(s, t)}$ | 31.18 | 31.39 | 31.54 | 31.73 | 31.87 | 32.03 |

TABLE 4.2: Estimates of the limiting SST upper bound in (4.15).

4.5. Discussion and Conclusion

Rather than removing trend and seasonal components before modeling residual excess above a threshold, trend and seasonal terms could have been included in the shape and scale parameters of the GP distribution. For example, a constant threshold u could have been used (or perhaps $u = u(s)$) along with a scale parameter of the form $\sigma = \sigma(s, t)$ that would have resulted in the same type of upper bound as in (4.14). A similar technique was used in Katz et al (2002) using the GEV distribution for precipitation maxima.

The SST upper bound estimates summarized in Table 4.2 are comparable to the 31.85°C proposed originally proposed in Ramanathan and Collins (1991). While this work provides evidence of the distribution of SSTs having a bounded tail, this analysis is done using a very limited number of years worth of SST data in the WPWP, and so any inference as to an SST upper bound is limited only to a small region of space and time.

Since Ramanathan and Collins, maximum SST values have been suggested from 27-32°C (Wilson and Opdyke (1996)), 28-32°C (Pearson et al (2001)), 30-33°C (Wilson et al (2002)), and 30.4°C (Li et al (2000)). However, temperatures derived from the paleontological record provide conflicting evidence of a thermostat. In a multiple proxy and model study of Cretaceous upper ocean temperatures and atmospheric CO₂ concentrations, Bice et al (2006) obtain temperatures between 33 and 42°C. The estimates obtained in this work are obtained using extreme value theory, are meant to compliment the others mentioned above.

There is room for improving the GP model proposed here for modeling SST extremes from which SST upper bound estimates are obtained. Incorporating cloud-SST feedback, evaporation-wind-SST feedback, and ocean dynamics and heat transport (the three SST-regulation mechanisms) into the model could result in a more accurate (albeit more complex) spatial model from which an estimated SST upper bound could be obtained. Longer SST time series along with measurements from more locations would also be beneficial.

5. GENERAL CONCLUSIONS

The extreme value index (EVI) ξ links the generalized extreme value (GEV) distribution (see (4.3)) and the generalized Pareto (GP) distribution (see (1.1)). These two distributions are fundamental in extreme value theory (EVT), with the GEV distribution being the only possible non-degenerate limiting distribution of properly normalized maxima of iid random variables, and the GP distribution appearing as the limit distribution of scaled excesses over high thresholds.

A new family of estimators for $\xi^{-1} = \alpha$, in the case $\xi > 0$, is given in (2.5). These estimators include the popular Hill estimator (see (2.3)), an estimator obtained by matching theoretical and empirical harmonic means ($\theta = 1$ in (2.5)), as well as more robust estimators of α (for example, (2.18)). Theoretical properties such as convergence in probability (Theorem 2.2.1.1), asymptotic normality (Theorem 2.2.1.2), asymptotic contamination breakdown point (Theorem 2.2.2.1), and gross error sensitivity (Theorem 2.2.2.2) are developed in Chapter 2. The usefulness of the new harmonic moment estimators are highlighted in Sections 2.3-2.5 through simulations and applications.

An estimator for $\xi^{-1} = \alpha > 0$ is given in (3.9) in the case one is working with grouped data. Existence and uniqueness of the estimator is given in Lemma 3.3.0.1. The focus of Chapter 3 is on using this estimator in insurance applications when exact claim amounts are unknown, but the number of claim amounts within certain intervals is known. The

proposed estimator in this chapter is shown (Section 3.4) to be robust with respect to a large class of distributions commonly used in modeling large insurance losses, and compares favorably to the Hill estimator that uses individual data.

In Chapter 4 the thermostat hypothesis for sea surface temperature is considered by investigating at the sign of ξ . A GP model that includes latitude and time as covariates is used to obtain a sea surface temperature upper bound estimate of $31.2^\circ - 32.0^\circ\text{C}$ in the western Pacific warm pool. However, this result is based on a limited number of years of SST data and only six different spatial locations. The three SST-regulation mechanisms discussed are discussed in Section 4.1.1, which are not included in the GP model. The methodology used in this chapter has the advantage over some others in that the daily SST records were used, rather than monthly or yearly averages, where information on extremes is lost. Also, extreme value theory is used to obtain upper bound estimates, rather than regression or other methods that focus on expected values.

BIBLIOGRAPHY

1. Beirlant, J., and Teugels, J. L., (1992), Modeling large claims in non-life insurance, *Insurance: Mathematics and Economics*, 11, 17–29.
2. ———, and Guillou, A., (2001), Pareto Index Estimation Under Moderate Right Censoring, *Scandinavian Actuarial Journal*, 2, 111–125.
3. ———, Matthys, G., and Dierckx, G., (2001), Heavy-Tailed Distributions and Rating, *Astin Bullitin*, 31(1), 37–58.
4. ———, Delafosse, E., Matthys, G., and Guillou, A. (2004), Estimating Catastrophic Quantile Levels for Heavy-Tailed Distributions, *Insurance: Mathematics and Economics*, 34(3), 517–537.
5. ———, Goegebeur, Y., Segers, J., and Teugels, J., (2004), *Statistics of Extremes*, Wiley, New York.
6. ———, Guillou, A., Dierckx, G., F-V., Amelie., (2007), Estimation of the extreme value index and extreme quantiles under random censoring, 10(3), 151–174.
7. Berkelmans, R., De'ath, G., Kininmonth, S., and Skirving, W.J., (2004), A comparison of the 1998 and 2002 coral bleaching events on the Great Barrier Reef: spatial correlation, patterns and predictions, *Coral Reefs*, 23, 7483.
8. Bianchi, K. L., and Meerschaert M. M., (2000), Scale and Shift Invariant Estimators for the Heavy Tail Index Alpha, *AMS 2000 Subject Classification: 62F12, 60F05*
9. Bice, K. L., D. Birgel, P. A. Meyers, K. A. Dahl, K. Hinrichs, and R. D. Norris (2006), A multiple proxy and model study of Cretaceous upper ocean temperatures and atmospheric CO₂ concentrations, *Paleoceanography*, 21, PA2002.
10. Brazauskas, V., and Serfling, R., (2000), Robust and Efficient Estimation of the Tail Index of a Single-Parameter Pareto Distribution, *North American Actuarial Journal*, 4(4), 12–27.
11. Brown, B.G. and Katz, R.W., (1995), Regional analysis of temperature extremes: Spatial analog for climate change?, *Journal of Climate*, 8, 108–119.
12. Buishand, T.A., (1991), Extreme rainfall estimation by combining data from several sites, *Hydrological Sciences Journal*, 36(4), 345365.
13. Caers, J., Beirlant, J., and Maes, M. A., (1999), Statistics for Modelling Heavy Tailed Distributions in Geology: Part II. Applications, *Mathematical Geology*, 31(4), 411–434.

14. Cebrián, A. C., Denuit, M., and Lambert, P., (2003), Generalized Pareto Fit to the Society of Actuaries' Large Claims Database, *North American Actuarial Journal*, 7, 18–36.
15. Coles, S., (2001), *An introduction to statistical modeling of extreme values*, Springer.
16. Dekkers, A. L. M., and de Haan, L., (1989), On the Estimation of the Extreme Value Index and Large Quantile Estimation, *The Annals of Statistics*, 17, 1795–1832.
17. ——— and ———, (1993), Optimal choice of sample fraction in extreme value estimation, *Journal of Multivariate Analysis*, 47, 173–195.
18. De Sousa, B., and Michailidis, G., (2004), A Diagnostic Plot for Estimating the Tail Index of a Distribution, *Journal of Computational and Graphical Statistics*, 13(4), 974–1001.
19. Drees, H., (1998), Smooth Statistical Tail Functionals, *Scandinavian Journal of Statistics*, 25(1), 187–210.
20. ———, de Haan, L., and Resnick, S., (2000), How to make a Hill Plot, *Annals of Statistics*, 29, 266–294.
21. Dupuis, D. J., (1999), Exceedances over high thresholds: A guide to threshold selection, *Extremes*, 1, 251–261.
22. Embrechts, P., Klüppelberg, C., and Mikosch, T., (1997), *Modelling Extremal Events*, Berlin: Springer-Verlag.
23. ———, Resnick, S. I., and Samorodnitsky, G., (1999), Extreme Value Theory as a Risk Management Tool, *North American Actuarial Journal*, 3(2), 30–41.
24. Fawcett, L., Walshaw, D., (2007), Improved estimation for temporally clustered extremes, *Environmetrics*, 18(2), 173188.
25. Fialová, A., Jurečková, J., and Picek, J., (2004), Estimating Pareto Tail Index Based on Sample Means, *REVSTAT*, 2(1) 75–100.
26. Finkelstein, M., Tucker, H. G., and Veeh, J. A., (2006), Pareto Tail Index Estimation Revisited, *North American Actuarial Journal*, 10(1), 1–10.
27. Fisher, R. A., and Tippett, L.H.C., (1928), Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proc. Cambridge Philos. Soc.*, 24, 180–190.
28. Fraga Alves, M. I., (2001), A Location Invariant Hill-Type Estimator, *Extremes*, 4(3), 199–217.

29. Fréchet, M., (1927), Sur la loi de probabilité de l'écart maximum. *Ann Soc. Math. Polon*, 6, 93–116.
30. Frigessi, A., Haug, O., and Rue, H., (2002), A Dynamic Mixture Model for Unsupervised Tail Estimation without Threshold Selection, *Extremes*, 5, 219–235.
31. Fu, R., Del Genio, A.D., Rossow, W.B., and Liu, W.T., (1992), Cirrus-cloud thermostat for tropical sea surface temperatures tested using satellite data, *Nature*, 358, 394–397.
32. Gilleland, E., and Katz, R.W., (2005), Extremes Toolkit: Weather and Climate Applications of Extreme Value Statistics, <http://www.isse.ucar.edu/extremevalues/tutorial/>.
33. Gnedenko, B.V., (1943), Sur la distribution limite du terme maximum d'une série aléatoire, *Annals of Mathematics*, 44, 423–453.
34. Gomes, M. I., Martins, M. J., and Neves, M., (2000), Alternatives to a Semi-Parametric Estimator of Parameters of Rare Events - The Jackknife Methodology, *Extremes*, 3(3), 207–229.
35. ———, Castro, L. C., Fraga Alves, M., and Pestana, D., (2007), Statistics of extremes for IID data and breakthroughs in the estimation of the extreme value index: Laurens de Haan leading contributions, *Extremes*, 10(6).
36. Groeneboom, P., Lopuhaä, H. P., and de Wolf, P. P., (2003), Kernel-Type Estimators for the Extreme Value Index, *The Annals of Statistics* 31(6), 1956–1995.
37. Hall, P., (1990), Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems, *Journal of Multivariate Analysis*, 32, 177–203.
38. Henry III, J. B., and Hsieh, P.-H., (submitted), Extreme Value Analysis for Partitioned Insurance Losses.
39. Hill, B. M., (1975), A Simple General Approach to Inference about the Tail of a Distribution, *The Annals of Statistics*, 3, 1163–1174.
40. Hoegh-Guldberg, O., Mumby, P.J., Hooten, A.J., Steneck, R.S., Greenfield, P., Gomez, E., Harvell, C.D., Sale, P.F., Edwards, A.J., Caldeira, K., Knowlton, N., Eakin, C.M., Iglesias-Prieto, R., Muthiga, N., Bradbury, R.H., Dubi, A., and Hatziolos M.E., (2007), Coral Reefs Under Rapid Climate Change and Ocean Acidification, *Science*, 318, 1737–1742.
41. Hogg, R. V., and Klugman, S. A., (1984), *Loss Distributions*, New York: Wiley.

42. Hosking, R.M., Wallis, J.R., and Wood, E.F., (1985), Estimation of the generalised extreme-value distribution by the method of probability-weighted moments, *Technometrics*, 27, 251-261.
43. Hsieh, P.-H., (1999), Robustness of Tail Index Estimation, *Journal of Computational and Graphical Statistics*, 8(2), 318–332.
44. ———, (2002), An Exploratory First Step in Teletraffic Data Modeling: Evaluation of Long-Run Performance of Parameter Estimators, *Computational Statistics and Data Analysis*, 40, 263–283.
45. Juarez, S., and Schucany, W., (2004) Robust and Efficient Estimation for the Generalized Pareto Distribution, *Extremes*, 7, 237–251.
46. Katz, R., Brush, G., and Parlange, M., (2005), Statistics of extremes: Modeling ecological disturbances, *Ecology*, 86, 1124-1134.
47. ———, Parlange, M., and Naveau, P., (2002), Statistics of extremes in hydrology, *Advances in Water Resources*, 25, 1287–1304.
48. Kleypas, J.A., (2006), Predictions of climate change in the tropical oceans, and how that should shape reef conservation efforts, *Proceedings of the World Marine Technology Conference*, London, Mar 2006, 8pp (CD-ROM only, ISBN no. 1-902536-54-1).
49. ———, G. Danabasoglu, and J. M. Lough (2008), Potential role of the ocean thermostat in determining regional differences in coral reef bleaching events, *Geophys. Res. Lett.*, 35.
50. Li, T., Hogan, T.F., and Chang, C.P., (2000), Dynamic and Thermodynamic Regulation of Ocean Warming, *Journal of the Atmospheric Sciences*, 57, 3353–3365.
51. Pearson, P.N., Ditchfield, P.W., Singano, J., Harcourt-Brown, K.G., Nicholas, C.J., Olsson, R.K., Shackleton, N.J., and Hall, M.A., (2001), Warm tropical sea surface temperatures in the Late Cretaceous and Eocene epochs, *Nature*, 413, 481–487.
52. Markovich, N., (2007), *Nonparametric Analysis of Univariate Heavy-Tailed Data: Research and Practice*, Wiley.
53. Maronna, R. A., Martin, R. D., and Yohai, V. J., (2006), *Robust Statistics: Theory and Methods*, Wiley.
54. McNeil, A. J., (1997), Estimating the Tails of Loss Severity Distributions Using Extreme Value Theory, *ASTIN Bulletin*, 27, 117–137.
55. Peng, L., and Welsh, A. H., (2001), Robust Estimation of the Generalized Pareto Distribution, *Extremes* 4(1), 53–65.

56. Pickands, J. III, (1975), Statistical Inference Using Extreme Order Statistics, *The Annals of Statistics*, 25, 1805–1869.
57. Ramanathan, V., and Collins, W., (1991), Thermodynamic regulation of ocean warming by cirrus clouds deduced from observations of the 1987 El Niño, *Nature*, 351, 27–32.
58. Renyi, A., (1953), On the Theory of Order Statistics, *ACTA Mathematica Academiae Scientiarum Hungaricae*, 4, 191–232.
59. Ribatet, M. A., (2006), A User’s Guide to the POT Package (Version 1.4), <http://cran.r-project.org/>.
60. Richardson, A. J., and Schoeman, D. S., (2004), Climate Impact on Plankton Ecosystems in the Northeast Atlantic, *Science*, 305, 1609–1612.
61. Segers, J., (2005), Generalized Pickands estimators for the extreme value index, *Journal of Statistical Planning and Inference*, 128(2), 381–396.
62. Smith, J.A., (1989), Regional Flood Frequency Analysis Using Extreme Order Statistics of the Annual Peak Record, *Water Resources Research*, 25(2), 311–317.
63. Smith, R.L., (1985), Maximum likelihood estimation in a class of nonregular cases, *Biometrika*, 72, 6790.
64. ———, (1999), Trends in rainfall extremes, (unpublished paper) available at http://www.stat.unc.edu/faculty/rs/papers/RLS_Papers.html.
65. von Mises, R., (1936), La distribution de la plus grande de n valeurs. *Rev. Math. Union Interbalcanique*, 1, 141–160.
66. Wallace, J.M., (1992), Effect of deep convection on the regulation of tropical sea-surface temperature, *Nature*, 357, 230–231.
67. Weare, B.C., Strub, P.T., and Samuel, M.D., (1981), Annual Mean Surface Heat Fluxes in the Tropical Pacific Ocean, *Journal of Physical Oceanography*, 11, 705–717.
68. Wilson, P.A., and Opdyke, B.N., (1996), Equatorial sea-surface temperatures for the Maastrichtian revealed through remarkable preservation of metastable carbonate, *Geology*, 24(6), 555–558.
69. ———, Norris, R.D., and Cooper, M.J., (2002), Testing the Cretaceous greenhouse hypothesis using glassy foraminiferal calcite from the core of the Turonian tropics on Demerara Rise, *Geology*, 30(607), 607–610.